ESSAYS IN SERVICES MARKETING: CAPACITY, PRICING, COMPLEMENTARITY, COMPETITION AND SEASONALITY

By

HYONGJAE RHEE

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1997

ACKNOWLEDGMENTS

First of all, I am immensely indebted to my Lord Jesus Christ for His support and guidance in my research. I also believe that God would protect and guide my life in the future as before.

I would like to thank deeply Professor Steve Shugan. He is the best advisor from whom I have ever received the academic guidance. He has consistently supported and encouraged my research through my doctoral study. Dr. Shugan has spent his precious time discussing my research with me. Discussion with him was surely a fun as well as an exciting intellectual exercise. I greatly benefited from his way of creative thinking. Steve Shugan will be in my heart as a typical prominent scholar.

I would also like to thank Professor Barton Weitz for his interesting comments and warm support. I appreciate Professor Murali Mantrala's detailed comments on this thesis. Professor Jayashree Mahajan has provided me with helpful comments. As an external committee member, Dr. Sanford Berg has helped my research from an economic perspective. I wish to thank the excellent faculty of the Marketing Department for their kind support. In particular, I have really enjoyed the lectures by Professors Steve Shugan, John Lynch, Bart Weitz and Richard Lutz.

Finally I wish to thank my parents, Hyun-Sik Rhee and Hye-Won Kim, for their continuous financial support and mental encouragement. I would also like to thank my

wife, Yoon-Jung, for her warm support and good care of our two daughters, Jung-Won and Jung-Bin. Our children have been one of the great sources of my delight.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

ABSTRACT vi

CHAPTERS

1 INTRODUCTION 1

2 SERVICES PRICING STRATEGIES: DEALING WITH CAPACITY USE 6

- 2.1 Introduction 6
  - 2.1.1 Peak Load Pricing Models 6
  - 2.1.2 Applications to Services Pricing 10
- 2.2 Peak Load Pricing in Marketing 13
- 2.3 Model 16
- 2.4 General Implications 22
- 2.5 Specific Implications 23
- 2.6 Empirical Resaerch 26
- 2.7 Conclusions and Future Research Marketing 28

3 A REVIEW OF SERVICES COMPETITION IN CAPACITY AND PRICE 32

- 3.1 Framework of Review on Services Competition 32
  - 3.1.1 Structure of Services Competition 32
  - 3.1.2 Role of Capacity Constraints 33
  - 3.1.3 Proposed Framework 34
- 3.2 Price Competition with Capacity Constraints 39
  - 3.2.1 A Simple One-Stage Pricing Game with Fixed Capacity 40
  - 3.2.2 Specifications of Residual or Contingent Demand 41
  - 3.2.3 Basic Models 44
  - 3.2.4 Service Reliability 47
- 3.3 Capacity Choice and Pricing in Services Competition 51
  - 3.3.1 Basic Models 52
  - 3.3.2 Service Reliability 59
  - 3.3.3 Switching Costs 64

3.4 Implications from the Literature 70
    3.4.1 Price Dispersion: Good or Bad 70
    3.4.2 Excess Capacity in Equilibrium 73
3.5 Conclusions and Future Research Directions 74

4 COMPETITIVE CAPACITY DECISION AND PRICING IN SERVICES 80

4.1 Introduction 80
4.2 Model Formulation 82
4.3 Service Capacity Decision 87
4.4 Service Pricing Decision 90
4.5 Effects of Demand Fluctuations 93
4.6 Empirical Analysis and Managerial Implications 96
4.7 Summary and Directions for Future Research 100

APPENDICES

A. SERVICES PRICING EXAMPLES 103
B. PROOFS OF PROPOSITIONS IN CHAPTER 2 105
C. PROOFS OF PROPOSITIONS IN CHAPTER 4 107

REFERENCES 110

BIOGRAPHICAL SKETCH 116

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ESSAYS IN SERVICES MARKETING: CAPACITY, PRICING, COMPLEMENTARITY, COMPETITION AND SEASONALITY

By

Hyongjae Rhee

December 1997

Chairman: Dr. Steven M. Shugan
Major Department: Marketing

The purpose of this dissertation is to investigate primary issues in services marketing. The topic of chapter 2 is services pricing strategies: dealing with customer capacity use. Demand for service fluctuates in a predictable pattern: high season and low season. Since services can not be stored, unsold service capacity has no value for service firms. Further capacity requires a huge amount of capital and is fixed in the short run. Service firms' profitability may depend on how to make efficient use of capacity. The proposed model for the multi-service firm suggests that pricing, given a capacity constraint, should depend on the two dimensions: type of services and extent of capacity use. Empirical research has been done through collecting data in the restaurant industry. It supports a key implication of the model that services taking capacity should have larger markups than services taking no capacity, all else being equal.

Chapter 3 reviews and proposes a framework to find the optimal capacity and pricing strategies in the competitive services market. The framework has three dimensions: reliability or availability, search and switching costs. Service reliability or availability is a supply-side variable and search is demand-side dimension while switching costs are due to a mixture of supply- and demand-side factors. This chapter organizes the literature on service competition via the proposed framework. The literature review provides two important managerial implications: multiple prices for the same kind of service exist and excess capacity exists in equilibrium.

The final chapter develops and analyzes a comprehensive model of capacity and pricing decision-making in a competitive market, allowing for services demand to fluctuate randomly as well as seasonally. The proposed model investigates a two-stage sequential process where capacity decision is made first and then optimal pricing is determined. The model predicts that capacity decision should depend on degree of competition, degree of demand fluctuations, inter-period price sensitivities. The pricing strategies should also depend on degree of competition, price sensitivities and capacity levels. The validity of the analytical results is checked by an empirical analysis using the data in the hotel industry that come from the published sources.

CHAPTER 1
INTRODUCTION

Demand for service fluctuates over time as demand for manufactured goods does. Besides random variations, service demand has predictable changes such as high-season and low-season demand. Since services are, in nature, not storable for sale in the next period, unused or unsold service capacities have no value for firms at all (Png 1989; Deneckere and Peck 1995). For instance, this phenomenon can be observed in empty seats in a flight, unoccupied hotel rooms, non-rented cars and so on. On the other hand, when demand exceeds a firm's capacity, the firm will lose excess demand to its competitors. Hence service capacity is a primary strategic consideration (Shugan, forthcoming) especially when firms consider investing a large amount of capital to build capacities to serve visiting customers. Further, unlike other marketing mix variables, firms have limited flexibility in the short-term to adjust capacities to demand.

In chapter 2, we consider a situation where service firms face a pricing problem because of capacity constraints. In other words, service pricing is limited to the short run where service capacity may be fixed. As an extension of the existing peak load pricing framework where utilities charge different prices for shifting load from peak to off-peak, multi-service firm problems are investigated with the assumption that the service firm offers two kinds of services in the service line: mutual substitutes or complements. In restaurants, one entrée is a substitute for another entrée while one appetizer or one dessert is a complement for the entrée. In the multi-service pricing problem, demand for one service in peak period does not depend only on off-peak prices of the same service but also on peak and off-peak prices of the other service. However, the effect of competitive pricing is not considered here in order to focus attention on effect of service type on pricing with capacity constraints. The proposed model suggests that given capacity constraints, optimal peak and off-peak price should depend on marginal costs and demand characteristics such as own price elasticities and cross price elasticities. Given the same marginal costs and same price elasticities, optimal service prices should depend on service type and extent of capacity use. With all else being equal, substitute service consuming capacity should have a highest price and complementary service taking no capacity should have a lowest price, while a substitute consuming no capacity and a complement taking capacity should be priced in the middle. The proposed model suggests that multi-service pricing with capacity constraints depends on the type of services and extent of their taking capacity. Empirical research is also done in a real restaurant situation for verifying the key analytical results: a service taking capacity should have a higher mark-up than a service taking no capacity.

In chapter 3, we review the previous literature on competition in services and the implications for capacity and pricing decision-making. A critical review of the two streams of research on service competition (namely, an one-stage model of price competition with capacity constraints and a two-stage model involving both capacity decision and price competition) is facilitated by analyzing the work on following dimensions: service reliability, customer search and switching costs. Service reliability is closely related to firms' decisions on capacity and operation; thus this is a supply-side dimension. Customer search is a demand-side dimension in that it depends on degree of customer expertise, attitude toward risks and costs of search. Finally, switching costs can be customer-specific like learning costs as well as firm specific, e.g., many service firms strategically manage switching costs to increase their loyal customer base. Hence the dimension of switching costs is a mixture of demand-side and supply-side factors. One implication that can be drawn from the literature is that there exist multiple prices for the same kind of service in equilibrium. Out of the factors that make multiple prices exist for the same service, price dispersion is good for customers when they are heterogeneous in preferences on a combination of price and service reliability (or service availability.) For instance, in order to take auto repair service, those who have high preferences for service reliability or availability, pay high prices and get the repair service at a reserved time while those who have low preferences, pay low prices and get the service on a first-come first-serve basis. The reason why this kind of price dispersion desirable to consumers is that customers are allowed to self-select a combination from a menu of options so that their utilities are maximized. Price dispersion related to service reliability or availability is unique to services since services cannot be inventoried. Demand uncertainty also results in price dispersion. In the presence of demand uncertainty, in order to avoid service outage, firms should have excess capacities that create additional costs and thus damage profitability. Loyal customers who purchase service regularly may be allowed to pay lower prices than irregular buyers. Likewise those who buy the service in advance of actual time of service can allow firms to reduce excess capacity should be able to pay lower prices than those who do not. For instance, subscribers to a magazine can get lower rates for each issue than non-subscribers do. When consumers do not have perfect information on price distribution for a service in the market, firms may be able to price discriminate customers who have varying degree of imperfect information. High search costs and thereby limited customer search may also turn firms into local monopolists who are able to price discriminate. In this case, price dispersion is not advantageous to customers since the "monopolists" would be able to extract consumer surplus. Another implication from the literature is excess capacity in equilibrium. Demand uncertainty poses need for excess capacity to firms in particular when service denial due to lack of enough capacity could have a substantially detrimental impact on firms' profitability; e.g., permanent loss of business. Excess capacity may also deter new entry into the market and also preempt incumbent competitors trying to deviate from the equilibrium prices. In other words, excess capacity acts as a strategic threat to possible new entrants and competitive incumbents.

In chapter 4, we analyze a comprehensive model of capacity decision and pricing in both deterministic and stochastic demand situations. Although there might exist multiple periods in terms of magnitude of demand, two periods are considered, for simplicity: season and off-season. In a monopolistic case, seasonal demand is assumed to hit the capacity and excess demand may be denied for service. For instance, tickets for a big event such as the Super Bowl may be sold out so that some people might not be able to get tickets unless they pay much higher prices to re-sellers. No resale is, however, assumed in the model. It is assumed that service firms have enough demand during high-season to fill up the capacity while demand for each firm is way below capacity in a low-season period. Further, in the competitive situation, services by one firm is, to some extent, differentiated with competitors' services in the sense that services are hard to standardize so there might be lot of variations. Each service firm competes in capacity decision and pricing with differentiated services that are not perfect substitutes. The way we set up the model is a two-stage process. Optimal capacity is determined in the first stage and then in the next stage optimal high-season and low-season prices are set given the capacity. This sequential model is realistic in the service industries since service firms may build capacities in the beginning and keep them in the short run, and then the firms may be able to adjust prices for matching the capacities to demand. The proposed sequential model is solved via a backward induction process. Thereby the equilibrium solutions are subgame perfect. Discount depth or difference between peak and off-peak prices is determined in equilibrium. In order to obtain managerial implications for service managers, we will make comparative statics by differentiating decision variables with respect to parameters such as degree of competition or substitutibility, capacity costs, and high season and low-season price sensitivities. The proposed model predicts that service capacity is negatively related with cross-period price sensitivity and with degree of demand fluctuation while the capacity is positively related with degree of competition. Regarding pricing decisions, the model predicts that both high-season price and low-season price increase as capacity level gets larger. As cross-period price sensitivity gets larger, high-season price is predicted to decrease while low-season price is predicted to increase. The model also predicts that discount depth is positively related with degree of competition and that cross-period price sensitivity negatively affects discount depth. In order to check the validity of analytical solutions in real service businesses, data in the hotel industry have been collected. As a measure of competition, the number of hotels in one area is used. Variance of occupancy rates is used as a measure of demand fluctuations. The empirical findings are supportive of the proposed model with demand seasonality.

CHAPTER 2

SERVICES PRICING STRATEGIES: DEALING WITH CUSTOMER CAPACITY USE

2.1 Introduction

Like demand for manufactured goods, services demand fluctuates over time. Moreover demand for services also varies on a seasonal basis in a predictable pattern. Unlike manufactured goods, however, services are not storable for next period's sale. Thus unused or unsold service capacities in one period have no value for firms. It is not easy to change service capacity depending on demand in every period. Therefore services firms need adjust their production capacities to fluctuating demand by basing their pricing scheme on time-of-use (TOU). For instance, restaurants may serve a discounted early special dinner or breakfast specials for early birds; also happy hours at bars, family nights at ball parks on weekdays and twilight specials at golf clubs. Similar TOU pricing schemes can be observed in rates for long distance or international calls and car rental rates. We can also observe seasonal pricing examples such as air fares and hotel rates.

2.1.1 Peak Load Pricing Models

The TOU pricing scheme originally has been extensively and intensively discussed and developed in the context of peak-load pricing (PLP) in the economics literature. The theory of peak load pricing has focused on pricing rules for public utilities. The PLP problem arises when a utility's product is economically non-storable and demand varies over time. Under these situations, uneven use of capacity can happen. To solve the pricing dilemma, utilities turned to a TOU rate structure, charging relatively higher prices for peak usage and lower prices for off-peak usage. The use of PLP for discouraging consumption during peak time and encouraging off-peak consumption can lead to more efficient use of a reduced level of capacity. With demand for capacity (i.e., unit operating cost is subtracted vertically from demand curve), Steiner (1957) assumed two equal-length periods of a day and solved the PLP problem for constant operating/capacity cost per unit and a capacity constraint. Hirshleifer (1958) generalized Steiner's long-run solution by separating the short-run and long-run aspects of the problem. His conclusion is that efficient utility pricing at any point in time should be based upon short-run marginal cost because all the utility contracts are on short-run basis (i.e., customers can cancel out their demand instantly) and thus at any point the relevant costs are short-run costs of meeting such demands. The welfare motivation of PLP was discussed by Williamson (1966), who extended Steiner's framework to handle periods of unequal length and obtained an effective demand for capacity using a weighted average of peak and off-peak periods, instead of summing the periodic demand vertically as was done by Steiner and Hershleifer. Platt (1981) provided synthesis of the above traditional PLP models by developing Willamson's effective demand for capacity to include any number of periods, any duration (equal or unequal) and any level of demand or of costs.

The initial models of PLP made the following assumptions: (a) demand is constant within pricing periods; (b) demand in one period is independent of demand in another; (c) production takes place under constant marginal costs; (d) the length and the number of pricing periods are exogenously given. Under these assumptions, the social objective is to maximize social welfare, i.e., the sum of consumer and producer surpluses. Essentially a unit of capacity is justified if and only if it is warranted by demand in any period alone or it is warranted by the combined demand in two or more periods. In a firm-peak case, where full capacity is only used in the peak, the total justified capacity is marginal capacity demand in the peak period. This capacity would be warranted even if there is no demand in off-peak. Thus off-peak demand could be allowed without additional capacity as long as the demand covers the operating costs of producing the service. Hence optimal peak price is set equal to the sum of marginal operating costs and capacity costs, and off-peak price is set equal to marginal operating costs. In a shifting-peak case, where full capacity is used in both periods, optimal off-peak price is greater than marginal operating costs and consumers in both periods contribute to capacity costs. In summary, outputs and prices in equilibrium are based upon equating the relevant demand prices to relevant marginal costs that are broadly defined as marginal production costs plus imputed price or shadow price for binding capacity constraints.

As a deviation from the classical assumption of independent demand over periods, substitution of off-peak for peak demand and shift of demand during peak period might be one of the main reasons for adopting PLP in the utility industry. As Berg and Tschirhart (1988) pointed out, smoothing demand between peak-period and off-peak period should depend on their own elasticities and cross elasticities of demand. Even though interdependencies between peak demand and off-peak demand has received relatively little attention in the literature, there have been efforts to introduce the interdependent demand. Pressman (1970) allowed for cross price effect between peak demand and off-peak demand by introducing cross elasticity of demand. In the model with capacity constraint and no profit constraint, he confirmed the firm-peak and shifting-peak equilibria derived from the classical PLP models.

Taking both time-varying and interdependent demand into account, Crew and Kleindorfer (1986) investigated not only welfare-maximizing equilibria but also equilibria of a profit-maximizing monopolist. Both optimal peak and off-peak prices in the monopolist solution depend on price elasticity of demand, and they are higher than optimal peak and off-peak prices in the welfare-maximizing solution. Bergstrom and Mackie-Mason (1991) also introduced marginal rate of substitution between peak and off-peak demand, and analyzed magnitude of peak/off-peak prices and change in capacity as compared to the uniform pricing with flat rates. A change from uniform pricing to PLP results in lower prices in both peak and off-peak periods, if the elasticity of substitution is greater than the ratio of off-peak price to off-peak profit margin (i.e., off-peak markup). If the elasticity of substitution is less than the ratio, moving toward PLP leads to higher peak price and lower off-peak price than the uniform price.

A potentially important issue of PLP may be how different rating periods are determined. A rating period is defined as an interval of time within a day where a single price is charged. Since utilities or regulatory commissions have direct control over peak prices and off-peak prices, and period lengths, it is worthwhile to analyze choices of length of pricing periods and effect of such choices on the optimal pricing policy. Load-duration curves could be used to show the relationship between demand quantities and time periods. The length and the specification of pricing periods might affect the load-shape. Gallant and Koenker (1984) focused on an average consumer behavior on an average day and using a rate treatment experiment, analyzed the effect of length of peak pricing period on the net welfare gains achievable by TOU pricing for residential electricity. Burness and Patrick (1991) dealt with continuous and interdependent demand and found that a profit-maximizing monopolist will choose a longer peak length than is socially optimal. An interesting point made by the authors is that in a dynamic framework in which demand is only a function of time and thus there are no cross-elasticities of substitution between demand at different times, optimal number and position of rating periods can be determined.

2.1.2 Applications to Services Pricing

Service pricing is often more complex than pricing manufactured products. Pricing complexity often comes from the ability of service providers to monitor usage and it allows service providers to increase profits with creative price mechanisms (Shugan in press). Recent research has attempted to incorporate the principles of peak load pricing into pricing models in the area of private businesses such as restaurants, theaters, hotels and airlines. Even though these businesses are similar to public utilities in terms of predictable variation in demand, the competitive businesses try to maximize their own profits rather than social welfare. Nagle (1984) pointed out that the principles of PLP, which accommodate temporal variation in nonstorable demand, are not only applicable to large public enterprises but also to such private enterprises as hotels, restaurants, health clubs, airlines, movie theaters and to some extent most retail establishments.

It is worth taking into account the customers' costs associated with buying service: inconvenience cost of using services at a less preferred time and crowding cost of waiting in a long queue (Gerstner 1986). Gerstner employed the symmetric Bertrand-Nash game as the framework between customers. Under commitment (i.e., sellers are assumed to commit themselves at the outset to prices for both periods), the price of peak period is higher than the price of off-peak period and the price difference is larger when inconvenience cost is larger. In this case, more customers buy in peak period with higher price and less in off-peak period if the inter-temporal value differential, i.e., inconvenience cost, becomes larger relative to crowding cost. However, when customers have little or no preference for time of use, there is no difference in price over periods, i.e., uniform pricing. Under no-commitment (i.e., sellers can change their prices in the second period when it arrives), on the other hand, for a large inconvenience cost, peak price is higher than off-peak price. For a small inconvenience cost, peak price is lower than off-peak price. In case of no commitment, more customers buy in the peak period irrespective of magnitude of inconvenience cost.

Another approach may postulate that demand by a customer is positively related to quantities demanded by other consumers, i.e., social influence on demand and price (Becker 1991). That could explain why some restaurants, sporting events such as the World Series and Super Bowls, theaters and other activities are reluctant to raise prices in the presence of persistent excess demand. Sometimes restaurants, bars or theaters faced with excess demand do not expand seating capacity because customers are fickle and a booming business is very fragile. Advertising and promotion may be designed for the bandwagon effect so that higher consumption by those vulnerable to publicity campaigns stimulates the demand of others.

Existing literature on the peak load pricing, however, has focused on single product, i.e., single utility (e.g., electricity, gas, water, etc.) or single service, so the literature lacks pricing implications for multiservice firms that may carry complements and substitutes in the service line. The literature also lacks pricing implications of extent of capacity usage.

The objective of this paper is to provide a framework for analyzing pricing mechanisms and to develop pricing strategies for multi-service firms. In the service industry, the main reasons for adopting the peak load pricing seems to be to encourage additional consumption at off-peak time as well as to shift demand from peak to off-peak period because of the capacity constraint during the peak period as in the utility industry. In other words, by reducing prices at off-peak time, we could make more efficient use of the given capacity and bring more new customers to stores.

In service businesses, it is also interesting to note the capacity-consuming attribute, since opportunity cost of capacity is large in peak period but is negligible in off-peak period. For instance, some restaurants train their waiters and/or waitresses not to recommend beverages or desserts in peak period so that they can seat more customers for meals. The restaurants may, however, serve free desserts and/or coffee in off-peak period so that customers could occupy tables as long as they want. Thus rationing of demand between peak and off-peak period may occur.

The remainder of this paper is organized as follows. In the next section, we start with pricing examples in service businesses like restaurants, bars, movie theaters, golf clubs, car rental companies and so on. In fact, the price differences for service based on the time of use motivate this research for a systematic framework for pricing scheme in the service industry. In the model section, we develop a pricing model, which is characterized by a constrained optimization problem, and then we analyze optimal pricing strategies based on degree of capacity usage and demand characteristics. Conclusions and areas for future research are discussed in the final section.

2.2 Peak Load Pricing in Marketing

Service is defined as all non-goods producing industries while goods producing industries consist of manufacturing, agriculture, mining and the other extraction industries.[1] On the side of employment, service sector accounts for nearly three out of every four jobs,[2] and that number is expected to keep increasing.[3] On the side of production, the U.S. Bureau of Economic Analysis reported that service accounted for 67% of Gross National Product in 1995.

The peak load pricing which originated and has been developed within the domain of utilities in economics is prevalent in the area of service businesses. As in the utility industry, service is direct, not storable, so service firms cannot store extra service produced at one period in stock to provide in the future. Demand for service varies in a predictable pattern (e.g., restaurants and bars are crowded Friday nights, more customers go to movie theaters and golf clubs during weekends, hotels and airlines usually have higher demand during summer and so on). Since service cannot be transported, the consumer must be brought to the service delivery system with a given capacity. Balancing the supply and demand sides of a service industry is not easy. Managing capacity constraints may be the most critical issue facing services (Sasser 1976, Shugan in Press). Consequently services firms frequently adopt PLP to shift demand from peak period to off-peak period and they further make changes in service composition by bundling some services together to develop new demand during off-peak time. The followings are facts about how each service business prices its services depending on time of use. See Table A-1 in Appendix A.

---

[1] The U.S. government uses this definition.

[2] Data from the U.S. Bureau of Labor Statistics

[3] This trend is estimated by the Fortune Magazine

In restaurants, they serve special breakfast or dinner for early birds in off-peak period, in which desserts and/or beverages (sometimes with free refill) are offered for free, and prices are also reduced by a certain amount. Free desserts and free beverages could allow customers to occupy tables at off-peak for a longer period of time than at peak. But the extended occupancy does not cost restaurants much, because more than enough tables during off-peak time have little opportunity cost. Many bars provide happy hour specials in off-peak period. Some items are served, during happy hours, for reduced prices, and the 2 for 1 rule (i.e., buy one get one free) or even the 3 for 1 rule applies to draft, glasses of liquor and glasses of house wine.

Movie theaters usually charge varying admissions depending on movie showtimes. During the day, the regular prices are reduced for the matinee. On a specific day of the week, the admissions may be substantially decreased (e.g., every Wednesday is student day). Some theaters feature such special movies as movies for kids during a few days of the week to ease crowding in peak period. Similar circumstances can exist in golf clubs. After early afternoon, green fees (sometimes including cart) usually go down and further, after late afternoon, golf clubs may offer even smaller green fees.

Cars could be rented for varying rates during day of the week. Although the rates may differ depending on the size of car, car rental companies usually maintain steady price discrepancies between weekday rates and weekend rates over varying sizes of cars. Whereas weekday rates can be larger than weekend rates in business areas, weekend rates can be larger than weekday rates in entertainment areas. Weekly rates may be attractive in cases of car rentals of more than five days, average rate per day calculated from weekly rate is significantly smaller than weekday rate.

Telephone companies seem to use PLP most systematically, because they charge different rates for long-distance calls and overseas calls at different times of the day and on different days of the week. For long-distance calls, there are typically three tiers of rates: 8 am - 5 p.m., 5 p.m. - 11 p.m., and 11 p.m. - 8 am and weekend (i.e., 11 p.m. Friday - 5 p.m. Sunday). Although the three-tier rule may also apply to international calls, time of tiers depends on the country to which a call is made.

Seasonal peak load pricing can be found both in hotels and airlines. The rates offered by hotels of even same quality in same period, may vary between business areas and entertainment areas, since peak periods are weekdays in business areas but they are weekends in entertainment areas. Among entertainment places, hotel rates could differ depending on types of fun (e.g., peak season of beaches is summer and peak season of ski resort is winter). While hotels announce their varying rates for seasons of year in the beginning, airlines often change air fares very frequently without prior notice.

2.3 Model

We consider two services in two different periods (i.e., off-peak period and peak period). This extends the typical single utility case in public economics literature. In addition to the fact that each service in peak period is a natural substitute to same service in off-peak period, the two services can be complements or substitutes to each other. In the situation of restaurants, one is an entree and the other is beverage in the case of complements, while one is an entree and the other is another entree in the case of substitutes. We also posit constant elasticity demand for each service in each period. The reason is that the constant elasticity demand function has been extensively applied to both empirical and theoretical research, and moreover the functional form can be estimated because it is intrinsically linear through logarithmic transformation (Shugan 1989). The symbol "≡" denotes "is defined as." For complements, pricing Strategies given Capacity Constraints

Notation

$m_{ij} \equiv$ market potential for *j*th service in *i*th period; $i=$ p(peak), o(off-peak), $j=1,2$

$\varepsilon_{ij} \equiv -(\partial q_i / \partial p_i)(p_i / q_i)$; own price elasticity of *j*th demand in *i*th period; $\varepsilon_{ij} > 1$

$\varepsilon_{ij.kl} \equiv -(\partial q_j / \partial p_i)(p_i / q_j)$;cross elasticity of *j*th demand in *i*th period with respect to *l*th price in *k*th period; $i \neq j$; $0 < \varepsilon_{ij.kl} < 1$

$c_{ij} \equiv$ marginal costs of each service in a period; $i=$ p(peak), o(off-peak), $j=1,2$

$K \equiv$ capacity

$k \equiv$ marginal capacity costs

We consider constant elasticity demand function in peak and off-peak periods.

Demand

$q_{p1} \equiv m_{p1} P_{p1}^{-\varepsilon p1} P_{p2}^{-\varepsilon p1,p2} P_{o1}^{\varepsilon p1,o1} P_{o2}^{\varepsilon p1,o2}$; peak demand of 1st service

$q_{p2} \equiv m_{p2} P_{p1}^{-\varepsilon p2,p1} P_{p2}^{-\varepsilon p2} P_{o1}^{\varepsilon p2,o1} P_{o2}^{\varepsilon p2,o2}$; peak demand of 2nd service

$q_{o1} \equiv m_{o1} P_{p1}^{\varepsilon o1,p1} P_{p2}^{\varepsilon o1,p2} P_{o1}^{-\varepsilon o1} P_{o2}^{-\varepsilon o1,o2}$; off-peak demand of 1st service

$q_{o2} \equiv m_{o2} P_{p1}^{\varepsilon o2,p1} P_{p2}^{\varepsilon o2,p2} P_{o1}^{-\varepsilon o2,o1} P_{o2}^{-\varepsilon o2}$; off-peak demand of 2nd service

Further we assume change in price of service i has a greater effect on its own demand than on demand for any other service and thus own price elasticity is larger than cross price elasticities. The (inverse) demand function of each product in one period has a functional form as follows.

$P_{p1} = P_{p1}(q_{p1}, q_{p2}, q_{o1}, q_{o2})$; peak inverse demand of 1st service

$P_{p2} = P_{p2}(q_{p1}, q_{p2}, q_{o1}, q_{o2})$; peak inverse demand of 2nd service

$P_{o1} = P_{o1}(q_{p1}, q_{p2}, q_{o1}, q_{o2})$; off-peak inverse demand of 1st service

$P_{o2} = P_{o2}(q_{p1}, q_{p2}, q_{o1}, q_{o2})$; off-peak inverse demand of 2nd service

For the purpose of calculation convenience, we use inverse demand functions hereafter. Each demand is dependent on its own price, price for the other service in the same period, prices for the same service and the other service in a different period , since all the four prices have direct or indirect effects on a demand in one period.

For costs associated with the two services in two periods, we assume a linear cost function.

$$C(q_{p1}, q_{o1}, q_{p2}, q_{o2}, K) = c_{p1} q_{p1} + c_{p2} q_{p2} + c_{o1} q_{o1} + c_{o2} q_{o2} + k K$$

Except for capacity cost, the other costs are assumed to be linear with the quantities produced. For instance, electricity charges, gas charges, raw material costs and payroll for cooks, waiters and/ or waitresses would be linearly related to demand for each service in each period. Under the above assumption, the monopolist maximizes the following total profits from the two services across off-peak and peak periods subject to capacity constraints.

$$\text{Max } \Pi = [P_{s1}q_{s1} + P_{s2}q_{s2}] + [P_{o1}q_{o1} + P_{o2}q_{o2}] - C(q_{p1}, q_{p2}, q_{o1}, q_{o2}, K)$$

$$\text{s.t.} \quad q_{p1} + q_{p2} \leq K$$

$$q_{o1} + q_{o2} \leq K$$

$$P_{ij}, q_{ij} \geq 0, K > 0 ; i = s,o, j=1,2.$$

The first constraint is on capacity in peak period and the second one is on capacity in off-peak period. At least the first constraint is very likely to be binding for some successful restaurants.

Since the problem for the monopolist has capacity constraints, we set up the Lagrangian function,

$$L(q, K, \lambda) = [(P_{p1}q_{p1} + P_{p2}q_{p2} + P_{o1}q_{o1} + P_{o2}q_{o2}) - C(q_{p1}, q_{p2}, q_{o1}, q_{o2}, K)]$$

$$+ \lambda_1(K - q_{p1} - q_{p2}) + \lambda_2(K - q_{o1} - q_{o2})$$

For the Lagrangian, the necessary and sufficient Kuhn-Tucker conditions for the global maximum to the objective function are

$$q_i^* \geq 0, \; Lq_j \geq 0, \; q_j^* Lq_i = 0, \; \text{for all } i = p1, p2, o1, o2$$

$$K^* \geq 0, \; L_K \geq 0, \; K^* L_K = 0,$$

$$\lambda_j^* \geq 0, \; L_j \geq 0, \; \lambda^* L_j = 0, \text{ for all } j=1,2$$

where $Lq_i = \partial L / \partial q_i$, $L_k = \partial L / \partial K$ and $L_{\lambda j} = \partial L / \partial \lambda_j$ (Hands 1991).

For each condition, we have equality if the corresponding variable satisfies the inequality and vice versa. Thus differentiating L with respect to q_{p1}, instead of P_{p1} and setting equal to zero yields

$$\frac{\partial L}{\partial q_{p1}} = \frac{\partial P_{p1}}{\partial q_{p1}} q_{p1} + P_{p1} + \frac{\partial P_{o1}}{\partial q_{p1}} q_{o1} + \frac{\partial P_{p2}}{\partial q_{p1}} q_{p2} + \frac{\partial P_{o2}}{\partial q_{p1}} q_{o2} - c_{p1} - \lambda_1 = 0$$

Dividing the equation by P_{p1} ($\neq 0$) yields

$$\frac{\partial P_{p1}}{\partial q_{p1}} \frac{q_{p1}}{P_{p1}} + 1 + \frac{\partial P_{o1}}{\partial q_{p1}} \frac{q_{o1}}{P_{p1}} + \frac{\partial P_{p2}}{\partial q_{p1}} \frac{q_{p2}}{P_{p1}} + \frac{\partial P_{o2}}{\partial q_{p1}} \frac{q_{o2}}{P_{p1}} - \frac{(c_{p1} + \lambda_1)}{P_{p1}} = 0$$

$\partial q_{ij} / \partial P_{kl} = \partial q_{kl} / \partial P_{ij}$, $i \neq k$ and $j \neq l$, by symmetry of cross partial derivatives. Cross effects, $\partial q_{ij} / \partial P_{kl}$, can be separated into a substitution effect and an income effect. If we assume symmetric substitution effects in which effect of *j*th price in period *i* on *l*th demand in period *k* is the same as effect of *k*th price in period *l* on *j*th demand in period *i*, we need only check income effects. If the budget is not rigid, then the income effect would be negligible or zero. Even with the assumption of a fixed budget, the cross partials can be symmetric if the income elasticities, $(dq_{ij} / dy)(y/q_{ij})$ where y is income level, for all goods for the same individual are identical and either the income distribution remain fixed or the income elasticities for all individuals are identical. See Pressman (1970) for details. Hence we get

$$\frac{\partial P_{p1}}{\partial q_{p1}} \frac{q_{p1}}{P_{p1}} + 1 + \frac{\partial P_{p1}}{\partial q_{o1}} \frac{q_{o1}}{P_{p1}} + \frac{\partial P_{p1}}{\partial q_{p2}} \frac{q_{p2}}{P_{p1}} + \frac{\partial P_{p1}}{\partial q_{o2}} \frac{q_{o2}}{P_{p1}} - \frac{(c_{p1} + \lambda_1)}{P_{p1}} = 0$$

In case of complementary services,

$$\mathrm{P}_{p1}=\frac{(c_{p1}+\lambda_1)}{1-\frac{1}{\varepsilon_{p1}}+\frac{1}{\varepsilon_{o1,\,p1}}-\frac{1}{\varepsilon_{p2,\,p1}}+\frac{1}{\varepsilon_{o2,\,p1}}},\ \mathrm{P}_{p2}=\frac{(c_{p2}+\lambda_{11})}{1-\frac{1}{\varepsilon_{p2}}+\frac{1}{\varepsilon_{p1,\,p2}}-\frac{1}{\varepsilon_{o1,\,p2}}+\frac{1}{\varepsilon_{o2,\,p2}}}$$

$$\mathrm{P}_{o1}=\frac{(c_{o1}+\lambda_2)}{1-\frac{1}{\varepsilon_{o1}}+\frac{1}{\varepsilon_{p1,\,o1}}+\frac{1}{\varepsilon_{p2,\,o1}}-\frac{1}{\varepsilon_{o2,\,o1}}}\ \text{and}\ \mathrm{P}_{o2}=\frac{(c_{o2}+\lambda_2)}{1-\frac{1}{\varepsilon_{o2}}+\frac{1}{\varepsilon_{p1,\,o2}}+\frac{1}{\varepsilon_{o1,\,o2}}-\frac{1}{\varepsilon_{p2,\,o2}}}.$$

Likewise, in case of substitute services,

$$\mathrm{P}_{p1}=\frac{(c_{p1}+\lambda_1)}{1-\frac{1}{\varepsilon_{p1}}+\frac{1}{\varepsilon_{o1,\,p1}}+\frac{1}{\varepsilon_{p2,\,p1}}-\frac{1}{\varepsilon_{o2,\,p1}}}$$ and by the same token, we can obtain P_{p2}, P_{o1} and P_{o2}.

The objective function is concave, all the constraint functions are quasi-convex and own price elasticity is greater than 1, so second order conditions are satisfied. Since own price elasticity and cross price elasticities are constant, P_{p1}^* is a function of its cost and shadow price. The optimal peak price of 1st service, P_{p1}^*, is clearly inversely related to its own price elasticity, ε_{p1}, so higher P_{p1}^* can be charged for inelastic patrons who insist on peak period. Since 1st service in peak period is a natural substitute for the same service in off-peak period, P_{p1}^* is positively related to cross elasticity $\varepsilon_{o1,p1}$. The P_{p1}^* in complements case where two kinds of services are complements to each other, would be smaller than in substitutes case, because effect of change in peak price of 1st service on demand for 2nd service is greater in the same peak period than in off-peak period (i.e., $\varepsilon_{p2,p1} > \varepsilon_{o2,p1}$) and thus net effect of change in price of 1st service on demand for 2nd service ($-1/\varepsilon_{p2,p1} + 1/\varepsilon_{o2,p1}$) is positive for two complementary goods, whereas the net effect ($1/\varepsilon_{p2,p1} - 1/\varepsilon_{o2,p1}$) is negative for two substitutes. The rationale for lower price in complements case might be that the monopolist, faced with a larger market potential combined, would try to sell more services and increase total profits by reducing price. Similar insights can be drawn from P_{o1}^* , P_{p2}^* and P_{o2}^* .

Regarding the capacity constraints, from $\partial L / \partial K = 0$, $\beta = \lambda_1 + \lambda_2$ and λ_1, λ_2 can be considered shadow price for additional capacity in peak period, off-peak period, respectively (i.e., λ's represent foregone profits from losing customers due to lack of capacity).When $\lambda_1 > 0$ and $\lambda_2 = 0$ (i.e., capacity constraint is only binding in peak period), peak period would be charged the whole marginal capacity cost, because off-peak period does not utilize the full capacity. The optimal prices are

$$P_{p1} = \frac{(c_{p1} + \beta)}{1 - \frac{1}{\varepsilon_{p1}} + \frac{1}{\varepsilon_{o1,\,p1}} - \frac{1}{\varepsilon_{p2,\,p1}} + \frac{1}{\varepsilon_{o2,\,p1}}}, \quad P_{p2} = \frac{(c_{p2} + \beta)}{1 - \frac{1}{\varepsilon_{p2}} + \frac{1}{\varepsilon_{p1,\,p2}} - \frac{1}{\varepsilon_{o1,\,p2}} + \frac{1}{\varepsilon_{o2,\,p2}}}$$

$$P_{o1} = \frac{c_{o1}}{1 - \frac{1}{\varepsilon_{o1}} + \frac{1}{\varepsilon_{p1,\,o1}} + \frac{1}{\varepsilon_{p2,\,o1}} - \frac{1}{\varepsilon_{o2,\,o1}}}, P_{o2} = \frac{c_{o2}}{1 - \frac{1}{\varepsilon_{o2}} + \frac{1}{\varepsilon_{p1,\,o2}} - \frac{1}{\varepsilon_{o1,\,o2}} + \frac{1}{\varepsilon_{p2,\,o2}}} .$$

When $\lambda_1 > 0$ and $\lambda_2 > 0$ (i.e., capacity constraint is binding in both periods), peak period would share marginal capacity cost with off-peak period (i.e., $\beta = \lambda_1 + \lambda_2$ where $\lambda_1 > 0$, $\lambda_2 > 0$). See the previous general form of optimal prices. When $\lambda_1 = \lambda_2 = 0$, there would be no binding capacity constraints in any period, so for optimal prices, β drops out from P_{p1}^* and P_{p2}^*. Even though this case appears over-investment in service capacity, some service firms might have to do business this way because demand may not only fluctuate seasonally but also randomly due to unexpected events .

2.4 General Implications

The proposed model suggests that optimal prices should depend on two dimensions: service type and capacity use. Each service can be a substitute or a complement in the framework of a multiservice firm. Service capacity is only meaningful when there is a excess demand over capacity. Consequently the dimension of capacity use can be separated into two situations: taking capacity in the presence of excess demand and taking no capacity in the presence of excess demand or having excess capacity whether or not a service takes capacity.

Table 2-1 Services Pricing Depending on Service Type and Capacity Use

	Taking Capacity	Taking No Capacity or w/Excess Capacity
Substitute	(1) + +	(2) + -
Complement	(3) + -	(4) - -

A point of the table is that all else being equal, service taking capacity in the presence of excess demand should be priced higher than service taking no capacity in the presence of excess demand, or service having excess capacity should be irrespective of service type. In each row of the table, the price of left cell is higher than that of right cell.

Another point is that given the extent of capacity use, price for substitute service should be higher than price for complementary service, all else being equal. In each column of the table, the price of top cell is higher than that of bottom cell. The following inequalities summarize the points of the table.

Price for Substitute taking Capacity 〉 Price for Complement taking Capacity 〉 Price for Complement taking No Capacity or w/Excess Capacity

Price for Substitute taking Capacity	〉	Price for Substitute taking No Capacity or w/Excess Capacity	〉	Price for Complement taking No Capacity or w/Excess Capacity

Thus substitute service taking capacity should be priced highest, and complementary service taking no capacity or having excess capacity should be priced lowest, while price for a substitute taking no capacity or having excess capacity and price for a complement taking capacity should be in between the two. An example of Cell (2) is that pizza restaurants may issue discount coupons for "carry out or delivery only." Cell (4) may be illustrated by the fact that golf carts are usually charged the same rates or minimally different rates over time periods because they don't take capacity and moreover speed up the plays. Another example of cell (4) is that beverages or desserts may be offered for free during early dinner specials.

2.5 Specific Implications

As a special case to the above general framework, we suppose that in the complements case, the first service consumes capacity and second service does not at all. Then q_{p2} , q_{o2} are assumed away in the first, second capacity constraint, respectively. With revised constraints, we obtain slightly different optimal prices.

$$P_{p1} = \frac{(c_{p1} + \lambda_1)}{1 - \frac{1}{\varepsilon_{p1}} + \frac{1}{\varepsilon_{o1,\,p1}} - \frac{1}{\varepsilon_{p2,\,p1}} + \frac{1}{\varepsilon_{o2,\,p1}}}, \quad P_{p2} = \frac{c_{p2}}{1 - \frac{1}{\varepsilon_{p2}} + \frac{1}{\varepsilon_{p1,\,p2}} - \frac{1}{\varepsilon_{o1,\,p2}} + \frac{1}{\varepsilon_{o2,\,p2}}}$$

$$P_{o1} = \frac{(c_{o1} + \lambda_2)}{1 - \frac{1}{\varepsilon_{o1}} + \frac{1}{\varepsilon_{p1.\,o1}} + \frac{1}{\varepsilon_{p2.\,o1}} - \frac{1}{\varepsilon_{o2.\,o1}}}, \quad P_{o2} = \frac{c_{o2}}{1 - \frac{1}{\varepsilon_{o2}} + \frac{1}{\varepsilon_{p1.\,o2}} + \frac{1}{\varepsilon_{o1.\,o2}} - \frac{1}{\varepsilon_{p2.\,o1}}}.$$

where $\lambda_1 = \beta$, if capacity constraint is binding in peak period

$\beta - \lambda_2$, if capacity constraint is binding both in peak and off-peak period

0, otherwise

$\lambda_2 = \beta$, if capacity constraint is binding in off-peak period

$\beta - \lambda_1$, if capacity constraint is binding both in peak and off-peak periods

0, otherwise

The capacity cost would be split between first service in peak period and first service in off-peak period if the capacity constraints are binding in both periods. The whole marginal capacity cost would be charged to first service in whatever period the capacity constraint is binding if there is excess demand in just one period. Relaxing the assumption of same marginal costs by allowing for different marginal costs between services, we use concept of markup, i.e., ratio of price to marginal cost and eliminate cost effect on price. Markup for first service is larger than that for second service, all things being equal. This leads to the following.

Proposition 2.1 *Given a binding capacity constraint, service consuming capacity should have larger markups than service which does not consume any capacity, all else being equal.*

See Appendix B for proof of proposition 2.1. Also confer Rhee and Shugan (1997) for a linear demand case. The relative markup, ratio of profit margin to price (i.e., $(P_i^* - c_i)/P_i^*$) for P_{p1}^* is larger than for P_{p2}^* in this special case. The proposition suggests that with a binding capacity constraint, larger markup be obtained from capacity consuming service, because marginal capacity cost would be included in the price of the service. Likewise service requiring capacity for a long time would have larger markups than does service requiring capacity for a short time.

It is also interesting to look at price difference between peak period and off-peak period. Let d_1 denote price difference of 1st product between periods.

$$d_1 = P_{p1} - P_{o1} = \frac{(c_{p1} + \lambda_1)}{1 - \frac{1}{\varepsilon_{p1}} + \frac{1}{\varepsilon_{o1,p1}} - \frac{1}{\varepsilon_{p2,p1}} + \frac{1}{\varepsilon_{o2,p1}}} - \frac{(c_{o1} + \lambda_2)}{1 - \frac{1}{\varepsilon_{o1}} + \frac{1}{\varepsilon_{p1,o1}} + \frac{1}{\varepsilon_{p2,o1}} - \frac{1}{\varepsilon_{o2,o1}}}$$

As price elasticity of peak demand, ε_{p1}, gets smaller, the price gap between the two periods gets larger. As price elasticity of off-peak demand, ε_{o1}, gets larger, the price gap also gets larger. This leads to the following.

Proposition 2.2 *All else being equal, as peak demand gets more inelastic or as off-peak demand gets more elastic, the difference between peak price and off-peak price becomes larger.*

See Appendix B for proof of proposition 2.2. This proposition suggests that if we assume the same costs and same corresponding cross elasticities between peak and off-peak periods, inelastic peak demand and/or elastic off-peak demand would increase the discrepancy between price and off-peak prices. Thus a service firm can price discriminate according to price elasticities of demand in peak and off-peak period; higher price can be charged to peak time patrons, while lower price can be charged to price elastic off-peak customers. However, if peak marginal cost equals off-peak marginal cost, and own price elasticities and corresponding cross elasticities are the same in both periods (i.e., $\varepsilon_{p1} = \varepsilon_{o1}$,$\varepsilon_{o1,p1} = \varepsilon_{o2,p2}$, $\varepsilon_{p2,p1} = \varepsilon_{p1,p2}$ and $\varepsilon_{o2,p1} = \varepsilon_{o1,p2}$), then there is no price difference, so uniform pricing over the two periods would result. Furthermore, price reversals (i.e., off-peak price is higher than peak price, see Bailey and White (1974)) may occur, if the off-peak demand is sufficiently more inelastic than peak demand; for instance, some people are not able to purchase service at off-peak due to constraints of work hours.

2.6 Empirical Research

The model predicts that in restaurants, desserts taking seating capacity should have larger markups than do appetizers which do not take capacity, because desserts prolong duration of a meal, while appetizers are served during the time when customers are already waiting for their entrees. Some restaurants may also provide early dinner specials with free dessert at off-peak although desserts take capacity and have a large opportunity cost of keeping next customers waiting for being seated and even possibly forcing some impatient customers to leave the restaurants and thus desserts are high markup orders by customers at peak time.

Table 2-2 can test an implication of proposition 2.1. The price data on appetizers and desserts was collected from restaurants in Gainesville, Florida. It is difficult to calculate costs of each restaurant dish because there are many ingredients in it. Thus as a reasonable proxy for real restaurant costs, the prices at supermarkets of each appetizer or dessert were checked. Five appetizers and five desserts were selected from the menu based on the availability of cost data. The prices were divided by the appropriate number of servings and then averaged for four supermarkets of Gainesville.

Table 2-2 Mark-ups for Appetizers and Desserts of Restaurants

		Average Price in Restaurants(1)	Average Price/ Serving in Supermarkets(2)	Mark-up*** (1) / (2)
Appetizers	Buffalo Wings	$4.32(0.52)*	$2.39(0.25)	1.81
	Shrimp Cocktail	4.87(0.11)	1.88(0.08)	2.59
	Eggroll	1.09(0.28)	0.89(0.07)	1.22
	French Onion Soup	3.25(NA)**	1.03(0.33)	3.15
	Vegetable Steak Soup	2.95(NA)	1.13(0.32)	2.61
			Mean Markup	2.28(0.76)
Desserts	Apple Pie	$3.25(NA)	$0.48(0.09)	6.77
	Cheesecake	3.04(0.14)	0.63(0.18)	4.83
	Cherry Cobbler	2.99(NA)	0.46(0.11)	6.5
	Carrot Cake	3.12(0.24)	0.36(0.04)	8.67
	Coffee	1.08(0.17)	0.09(0.03)	12.0
			Mean Markup	7.75(2.74)

Note

*Standard Deviation in Parenthesis

**NA denotes single sample

*** Markups for desserts are significantly higher than those for appetizers ($p < 0.01$).

For achieving as much comparability between costs as possible, all restaurants dishes were compared to their "ready to serve" counterparts in the supermarkets. In the case of shrimp, however, the prices of uncooked fresh shrimp were used for comparison. We did this because the size of cooked shrimp did not match the size of shrimp on the menu and prices of shrimp highly depend on size (i.e., counts per pound).

In a t-test, markups for desserts were significantly higher than markups for appetizers at significance level of 0.01. The empirical results show that with controlling for costs, capacity taking should result in a higher markup, all else being equal. If prices for cooked and ready-to-use shrimp on the menu had been used, markup for shrimp cocktail could have been smaller. As prices for single serve pack coffee were also used for convenience of calculating number of servings, use of price for coffee in a large volume like a can would have increased markup for coffee. Consequently if we take these two markups into consideration, we could have more significant results.

2.7 Conclusions and Areas for Future Research

This paper provides a framework for analyzing pricing mechanism in service businesses by starting from the peak-load pricing model in economics literature and further extending it into multiservice domain including complementary service and/or substitute service that a firm may carry in its service line. The analysis focuses on determining optimal prices for service businesses whose products are nonstorable and whose demand fluctuates systematically over time. As in the utility industries which has been usually analyzed in economics, the service industry also may face excess demand in peak period and need to smooth demand over periods. Thus service businesses would charge different prices based on time of use to shift demand from peak period to off-peak period and further to increase off-peak consumption.

In the proposed model of two services (complements or substitutes) and two periods (peak and off-peak period), optimal prices depend on marginal costs and demand characteristics such as own price elasticities and cross price elasticities. Given same marginal costs and same price elasticities, optimal service prices should depend on service type and extent of capacity use. All else being equal, substitute service taking capacity should have a highest price and complementary service taking no capacity or having excess capacity should have a lowest price, while a substitute taking no capacity or having excess capacity and a complement taking capacity should be priced in the middle. Relaxing the assumption of same marginal costs and allowing for different costs between services, we use concept of markups to find that a service requiring capacity should have higher markup than does a service requiring no capacity. Empirical research done in the situation of restaurants supported this implication by showing that markups for desserts were significantly higher than markups for appetizers. The rationale may be that desserts take seating capacity and prolong duration of a meal, while appetizers are served during the time customers are already waiting for their entrees. Regarding price difference over time periods, the difference between peak price and off-peak price gets larger as peak demand gets more inelastic and off-peak demand gets more elastic. Hence a service firm can price discriminate according to elasticities of demand in peak and off-peak period; a higher price can be charged to price inelastic peak time patrons, while lower price can be charged to price elastic off-peak customers. Although the empirical research was done in the situation of restaurants, the proposed model and its propositions can be applied to any kinds of service businesses such as bars, car rental companies, hotels, airlines, movie theaters, golf clubs and so on.

An immediate extension of this research is to consider peak load pricing in a competitive situation. Given capacity constraints, service firms may compete in two periods: peak and off-peak. Off-peak demand may be more price sensitive than peak demand. The firms may have different amounts of capacities. For simplicity, in a duopoly, one firm may implement peak load pricing while the other competitor sticks with uniform pricing. In another case of duopoly, both of the service firms use peak load pricing. This modeling might shed light on under what conditions on demand capacity service firms should adopt peak load pricing rather than uniform pricing over periods.

An interesting extension of the research could include waiting time of customers in service establishments. While firms certainly try to make most efficient use of capacity, customers positively or negatively value waiting time. In one sense, long waiting time could suggest that a product is popular and thus could signal quality of the product. In the other sense, however, long waiting time would cause inconvenience of delayed service and it could even lead to losing customers.

Bundling can be used together with the time of use pricing. Specifically in restaurants, for a regular menu, some dishes can be bundled together for attractive prices, but for a la carte menu, each dish is served and priced separately. The two kinds of menus may be combined with regular hours and special hours. The pricing scheme in service businesses could be richer by including bundling in the time of use pricing.

Finally, the lengths of peak and off-peak periods and the number of switch points of within a demand cycle (e.g., daily, weekly, quarterly, etc.) may play a crucial role in determining the extent of substitutability. If the peak period is very short, customers easily purchase the service in off-peak period in which price is lower. On the other hand, a very long peak period makes it hard for customers to substitute. Consequently, both pricing period lengths and the switch points may affect optimal prices and thus optimal profits.

CHAPTER 3
A REVIEW ON SERVICES COMPETITION IN CAPACITY AND PRICE

3.1 Framework of Review on the Service Competition in Capacity and Pricing Decision

Since services are, in general, not storable for sale of next period, unused or unsold service capacities have no value for firms at all (Sasser 1976; Png 1989; von Ungern-Sternberg 1991; Deneckere and Peck 1995). For instance, this phenomenon can be observed in empty seats in a flight, unoccupied hotel rooms, non-rented cars and so on. On the other hand, when demand exceeds a firm's capacity, the firm would lose excess demand to its competitors. Hence service capacity is one of primary strategic considerations in particular when firms should invest a large amount of capital in building capacities to serve visiting customers and thus they have limited flexibility in the short run to adjust capacities to demand, as compared with other marketing decision variables such as prices and advertising.

3.1.1 Structure of Service Competition

Competition can be defined as the process by which independent firms vie with each other for customers in a market (Weitz 1985). Services competition might be more complex than manufactured product competition, since the non-storability of services requires more precise matching demand and supply in services. In service industries, firms may be involved in competition in two scenarios. In one way, service firms can compete in pricing given fixed service capacity. This is the case especially when it is very hard to predict future demand of a certain service due to demand fluctuations. This competitive structure is similar to Bertrand conjectures where firms independently and simultaneously decide on price given their capacities. The other way to look at competition between firms is a two-stage process. In the first stage, firms decide independently and simultaneously capacity. In the next stage, each firm learns others' capacities and the firms compete for demand in the Bertrand way through price competition. This two-stage scenario is relevant in the long run where firms have enough time to raise fund for adjusting capacities to demand and thus the firms have flexibility to change their capacities as well as prices.

3.1.2 Role of Capacity Constraint

As Moorthy (1985) points out, given homogenous products and equal constant marginal costs for each firm, if there are no capacity constraints, each firm prices at the marginal cost and has no profits in the Bertrand Nash equilibrium (a Nash equilibrium is defined as a strategy profile in which each firm's strategy is a best response to the other firms' strategies (Kreps 1990). This is often called the Bertrand paradox. That is, a small increase in price by one of the firms with undifferentiated products or services would end up with zero demand for the firm due to price undercut by competitors. Thus all the identical firms should, in equilibrium, charge the competitive price that equals marginal variable cost and earn no profits. These general results may not hold in the framework of service competition because service firms might be constrained by their capacities and the capacity constraints can drive pricing strategies. If sellers are subject to capacity constraints, the Bertrand solution is no longer an equilibrium. Instead, as in the Edgeworth solution where each firm in a duopoly with homogeneous products is able to produce only less than demand at the competitive price, D(c), introducing capacity constraint can solve the Bertrand paradox. A certain firm that increases its price slightly could have a residual nonzero demand at a price greater than marginal cost and therefore makes positive profits. This suggests that in the presence of capacity constraints, the Edgeworth solution rather than the Bertrand solution is an equilibrium to price competition. Technically the reason why the Bertrand Nash equilibrium does not exist when firms are constrained by capacity, is that discontinuity in firms' profit functions arises when the prices of two or more firms are equal (i.e., given no capacity constraints, a firm offering an infinitesimal amount of price cut would capture entire market and other firms lose all the demand). Capacity constraints can be used to ensure existence of mixed strategy equilibrium,[1] if not necessarily pure strategy equilibrium, for pricing game (e.g., Peters 1984). Therefore in the service industry where capacity is one of main decision variables to firms, it would be very worthwhile to analyze service competition in pricing with capacity constraints through Bertrand-Edgeworth game theory dealing with both capacity and price.

3.1.3 Proposed Review Framework

When we make a critical review of the two streams of research on service competition (an one-stage process of price competition with capacity constraints and a two-stage process of capacity and price competition), it might be effective to analyze each portion of the literature on the following dimensions. One interesting dimension is *service reliability*. Service reliability refers to the degree to which customers are able to purchase requested service at a desired point in time. Service reliability becomes an important concern when there is an excess demand over capacity. The reliability problem can happen in the forms of service denial or waiting in a long line. Hence reliability is very closely related to firms' capacity decision (i.e., supply-side). The capacity decision is not obvious when demand is uncertain and fluctuates over time periods in a predictable pattern or in a random way since it is not easy to adjust capacity to varying demand without leading to excess capacity in periods of low demand. In other words, there is a trade-off between serving more customers at time of high demand and carrying excess capacity at time of low demand. Another relevant dimension is *customer search*. Since the search is associated with monetary costs and/or costs of time (i.e., demand-side), customers may search for a service firm that gives them the highest utility before they visit the firm. This prior (to visit to first best firm) search depends on the magnitude of search costs. Based on the prior search, customers choose a firm with a best combination of price and service reliability. On arrival, customers observe length of waiting queue and they may join the queue if the price plus cost of expected waiting time is less than their reservation price. Otherwise, they may search for another firm after they visit first best firm. The posterior (to the visit) search will continue until a full price, which includes price charged, expected search cost and expected cost of waiting time, does not exceed customers' reservation price. See De Vany (1976) and Davidson (1988). The stopping rule of posterior search for service is unique to services in the sense that expected waiting time may require further search. That is, when a customer visits a service firm and finds that he needs to wait more than he can, s/he will not join the waiting queue or s/he will join the queue and leave for

---

[1] Pure strategy is defined as a strategy that a player takes for sure, i.e., with probability one while mixed strategy refers to a set of strategies in which a player i takes a strategy s_i with probability

another firm. The posterior search is not usually observed in manufactured products because they can be stored in inventory for next periods in which consumers are able to purchase without waiting much. Even in search for quality in manufactured products, consumers may search for an alternative brand or model after visit of a store but they do not consider waiting time inside the store. As a third dimension of service competition, we can consider *switching costs* that refer to opportunity cost to switching from a current service provider to another. Switching costs do not include only physical benefits such as frequent flyer benefits and initial transactions fee saving by sticking to the existing service firms; these benefits are usually contrived strategically by firms for creating loyal clientele. Switching costs can also include nonphysical benefits to customers who keep visiting the same service firms.

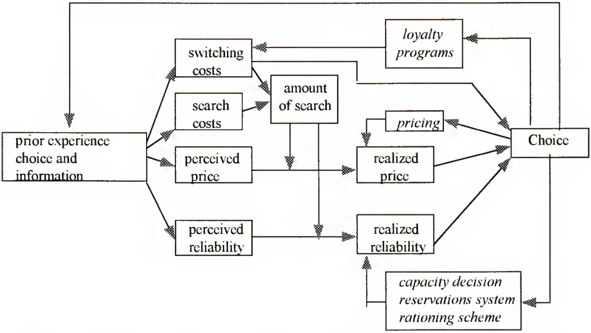


Figure 3-1-1 Diagram of Customer Decisions and Firm Decisions in a Service Industry

---

$p_i(s_i)$ where $\Sigma_i\, p_i(s_i)=1$ (Kreps 1990).

Those customers may save learning costs, in the forms of learning time and energy, that should be paid if they switch to another firm. This implies that nonphysical component of switching costs may be customer-specific. The switching costs then can be a demand-side as well as supply-side dimension. All the above three dimensions (service reliability, customer search and switching cost), in addition to price, can determine utility of a service. The diagram below summarizes services competition from customers' perspectives. The diagram shows how customers and services firms process information and make decisions. Managerial practices and policies of service firms are in italic letters.

Customer perception of a firm's service on price, reliability, search costs and switching costs are affected by prior purchase experiences and/or information from other customers. Considering costs of switching from the current firm and costs of search, customers decide amount of search, i.e., whether and how much they search for alternative firms.

The higher switching costs and search costs are, the lesser search customers would make. After search, price and reliability for service of each firm are realized. Firm pricing policies including promotion determines realized price. Firm capacity decision, reservations system and demand rationing scheme decide realized capacity. Reservations system is closely related to rationing since the system can control demand and ration excess demand over time. Customer switching costs are not only influenced by prior experiences (e.g., trust in a existing firm) but also by firm loyalty programs. Customers make service choice based on realized price, reliability and switching costs. Customer choices, as a feedback, influence firm strategies in pricing, capacity, reservation systems and schemes of rationing excess demand.

The following table is proposed to summarize a framework of service competition in capacity and price. The representative research is reviewed in sections 3.2 and 3.3.

Table 3-1-1 Framework for Reviewing Relevant Literature

	Price Competition given Capacity (Short-run)	Capacity and Price Competition (Long-run)
Basic Models	choice of optimal price with capacity constraint[(a)]	choice of optimal capacity and price[(b)]
Service Reliability (Supply-side)	(a) plus probability of being able to buy and service rate or waiting[(c)]	(b) plus (c)
Customer Search (Demand-side)	(a) plus customer search behavior[(d)]	(b) plus (d)
Switching Costs (Mix of Demand & Supply-side)	(a) plus endogenous or exogenous cost of switching to other firms[(e)]	(b) plus (e)

The two scenarios in the first row of the table denotes time horizons of analysis, in the short-run versus long-run. In the short run, service firms can only adjust prices to demand given fixed capacity while in the long run, the firms also have flexibility to change capacity, in addition to prices. Out of the three dimensions of the first column, service reliability associated with firms' capacity is supply-side factor while customer search that is related to customer characteristics such as prior knowledge on service, search costs, etc., is demand-side factor. Switching costs can, however, occur as learning cost and/or emotional attachment (i.e., customer-specific) or as lost benefit or initiation charge in case of switching (i.e., strategically designed by firms). Thus the switching costs are both demand-side and supply-side factors. Since all these three dimensions are relevant to services competition, analysis of services competition on the dimensions would be more realistic.

Non-price competitions in service industries are also feasible. Firms strategically may decide on their services on the dimension of customer tastes such as physical locations of service establishments. Service firms also make strategic decisions on quality of service in relation to quality levels of competitive services. The focus of the paper will be, however, on the literature of capacity-constrained price competition so that we should stay on the track of the main theme, services competitions based on capacity-constraints.

The chapter consists of five sections. The section two addresses price competition with capacity constraint on the dimensions of service reliability, customer search and switching costs. Customers decide amount of search for service firms other than the existing firm, depending on search costs and switching costs. Utility of each firm's service, as a function of price and service reliability, is evaluated for choice. In the subsection three, service capacity is not fixed and firms can choose both capacity and price as strategic decision variables. In the long run, it is more realistic that firms are flexible on capacity that is closely related to customer waiting time or service reliability. Once optimal capacity is determined, firms again compete in price. In the section four, we discuss the implications that are drawn from the literature. The final section summarizes the key points of the relevant literature and suggests directions for future research.

3.2 Price Competition with Capacity Constraints

In this section, we analyze a framework where service firms compete for customers by charging different prices given that the firms have fixed capacities. Prices are usually easier to adjust to demand, compared to capacity that takes substantial capital investment. The flexibility of price in relation to capacity may justify price competition given capacity constraints at least in the short run.

3.2.1 A Simple One-Stage Pricing Game with Fixed Capacity

Suppose there are two firms with rigid capacity constraints. The firms sell up to capacity. Suppose also that demand is certain at the moment. Demand uncertainty will be introduced in later subsections on service reliability. The demand function $P(q_1, q_2)$ is concave ($P'' \leq 0$) and the firms choose their prices simultaneously. Two cases are possible. One is a pure strategy equilibrium where the firms do not randomize in their choice of prices. Both firms charge the same price at which demand equals aggregate capacity in equilibrium. Such an equilibrium exists if and only if the capacities are not too high since lowering prices would not result in larger sales due to limited capacity. Therefore in that region of capacity space, the firms are engaged in Cournot game where they compete in quantities with the same market price. However, in the region of capacity that is greater than optimal capacity response to competitor's capacity, the firms randomize price within a certain range. In the interval, all prices that are chosen yield the same optimal payoff (for proof of these, see Tirole 1988).

When demand is easy to observe, larger capacity than optimal response to competitors' will be considered a burden to sell out for recovering investment rather than provide firms with competitive edge. In other words, large capacity is not cost effective under certain demand and high-capacity firm earns Stackleberg follower profits. On the other hand, in case that demand is uncertain and initial capacity choice is irrevocable, a larger firm with excess capacity would gain higher expected profits by providing a threat to its competitors and potential new entrants (Lieberman 1987;Benoit and Krishna 1987; Paraskevopoulos et al. 1991). Demand uncertainty or fluctuation may explain why service firms usually carry excess capacity at off-peak times.

3.2.2 Specifications of Residual or Contingent Demand

One of the key elements in models of capacity choice and subsequent price competition as well as capacity-constrained price competition is the specification of residual or contingent demand which is not satisfied by a low-price firm *j*. This situation implicitly assumes that firm j cannot meet all demand at its price, i.e., $D(p_j) > k_j$. The residual demand faced by firm *i*, min[k_i , max(0, $D(p_j)$ - k_j)], is a function of its own price and its competitors' prices, given their respective fixed capacities. Once capacities are fixed in the first stage, customers choose between firms that charge different prices in the second stage. Each customer first attempts to buy at low-price firm j and then unsuccessful customers try at high-price firm *i* as long as their reservation prices are greater than the high price. In this sequential demand behavior, customers may switch between firms without incurring any costs. There exist two possible interpretations of market demand. One is to assume that the market demand, $D(p)$, consists of identical individual consumers, all of whom share the same downward sloping demand curve. This is a bit too strong assumption, but consumer homogeneity is possible when a service firm focuses on a specific group of customers, e.g., college students in a college town. The other interpretation is that $D(p)$ is the summation of inelastic demands of heterogeneous consumers, all of whom want to purchase one unit of the service, as long as the price is not higher than their reservation price. When the low-price firm cannot meet all market demand, the sales of high-price firm depend on how the residual demand is rationed. Different rationing scheme would lead to different equilibria.

For the case of the first interpretation of market demand, Kreps and Scheinkman (1983) assume that low-price firm *j* realizes an excess demand for its product and places a limit, on the number of units that each customer may buy. A top portion of demand curve is satisfied and so sales for firm *i* will be $D(p_i)\text{-}k_j$. Hence its contingent demand curve is obtained by shifting the market demand curve that low-price firm *j* rations its available capacity efficiently in the second stage of price competition. The rationing rule supposes a residual demand function for high-price firm *i*:

$$\begin{array}{ll} D_e(p_i|\, p_j) = D(p_j)\text{-}\ k_j & \text{if } D(p_j) > k_j \\ \quad 0 & \quad \text{otherwise} \end{array}$$

Thus residual demand at the high-price firm *i* in a duopoly setting, equals market demand minus the capacity of the low-price firm *j*. Firm *i* always serves customers with higher valuation before customers with lower valuation. This scheme is obviously based on the assumption that firms can observe customer valuation for a given product or service. This rationing is called efficient because it maximizes consumer surplus in the sense that the low-price firm serves high-valuation customers.

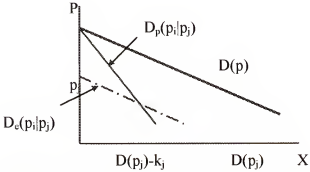


Figure 3-2-1 Rationing Schemes for Excess Demand

The efficient rationing which leads to the parallel shift in residual demand, denoted as $D_e(p_i|p_j)$ in the above graph. Thus residual demand at the high-price firm i in a duopoly setting, equals market demand minus the capacity of the low-price firm *j*. The results and their implications of efficient rationing will be discussed in section 2.1. Kreps and Scheinkman rationing rule may apply when service firms can know distribution of customer willingness to pay and serve first customers with higher reservation prices. It is also conceivable that if services firms have priority customers to serve, the efficient rationing would be appropriate since it allows firms to serve the priority customers before regular customers up to their capacities. For instance, in an auto service center, when service people are very busy, they might choose to serve first customers who ask for express oil change and are willing to pay higher prices.

Another approach to rationing residual demand starts from the notion that the efficient rationing scheme is too extreme. With identical consumers, the efficient rationing makes each customer's willingness to pay decrease equally for service of the high price firm. With consumer heterogeneity in reservation price, the parallel shift rationing by Kreps and Davidson allows the low-price firm j to skim most attractive customers from the market. In both interpretations of demand, the above rationing rule chosen by Kreps and Scheinkman (1983) can lead to the worst contingent demand for high-price firm i. Davidson and Deneckere (1986) suggest, by elaborating on Beckman (1965), an alternative rationing rule which leads to the best residual demand for firm i. This scheme is often called proportional rationing rule. Under first-come first-served rationing, the low-price firm j would not serve only some demand which would not be satisfied by the high-price firm i but would also serve some demand which can be served by the high-price firm i. Hence this rule is not efficient for customers. The contingent demand under the proportional rationing is denoted as

$$D_p(p_i|p_j) = D(p_i)\left[\frac{D(p_j) - k_j}{D(p_j)}\right]$$

However, the high-price firm i prefers this rationing rule to the efficient rationing rule, because its residual demand at its price p_i ($> p_j$), $D_p(p_i|p_j)$, is greater than the counterpart under efficient rationing scheme, $D_e(p_i|p_j)$. The alternative proportional rationing rule can be advocated in the following cases. If consumers are homogeneous in terms of reservation price, this rule rations output by allowing some customers to make unlimited purchase up to capacity. Equivalently, if the output is sold on a random basis like first-come first-served, when there are a large number of customers, the low-price firm serves a random sample of them.

This rationing rule may apply to services firms that do not usually take reservations or have priority customers to serve first. If customers have same willingness to pay or their distribution of reservation prices has a relatively small variation (e.g., fast-food restaurants), firms would rather serve customers on a first-come first-serve basis and the proportional rationing might be more relevant.

3.2.3 Basic Models

In this section, we review models that deal with price competitions given capacity constraints. These models, for the moment, do not consider product reliability, customer search or switching costs. In the models, residual demand is rationed in one of the two ways: efficient rationing rule or proportional rationing rule.

Suppose there are two firms in the market for simplicity. Firm i has capacity k_i ($k_1 \geq k_2 > 0$). Each firm produces the same service at the same constant marginal cost c. At the beginning of the game, each firm selects and announces a price which it will charge customers ($p_i < p_j$). The rationing rule of excess demand for each firm is fixed. Consumers first try to buy from firm i up to its capacity k_i. Then some of them who are not served turn to firm *j*. Although the rationing scheme is naturally chosen by firm *i*, it is not still certain which scheme is chosen since the rationing rule affects *j*'s profits, not *i*'s profits. According to nature of competition, we can assume, however, that firm *i* chooses the scheme which minimizes the profit of firm *j* since minimizing the other firm's profits means maximizing its own profits in the duopoly. For the purpose of discussion, we assume the efficient rationing scheme where those who have higher valuations of service are served first by the low price firm. Under these assumptions, the profits of firm *i* are

$$\begin{aligned} \Pi_i(p_i, p_j) &= p_i \min(k_i, D(p_i)) && \text{if } p_i < p_j \\ & p \min(k_i, k_i D(p)/ k) && \text{if } p_i = p_j = p \\ & p_i \min(k_i, \max(0, D(p_i) - k_j)) && \text{if } p_i > p_j \end{aligned}$$

where k is total capacity, i.e., $k = k_i + k_j$ and if $p_i = p_j$, demand is allocated in proportion to capacities.

In a Nash game which Osborne and Pitchik(1986) describes, firm *i* tries to maximize $\Pi_i(p_i, p_j)$ given k_i and k_j. There are three cases in terms of demand size relative to the given capacities of the two firms. In case that each firm has more than enough capacity to meet demand even at price equal to marginal cost, the capacity constraints are meaningless and there is a pure strategy equilibrium, i.e., $p_i = p_j = c$ (marginal cost) as in the Bertrand model. In case of excess demand in the industry, there exists also a pure strategy equilibrium since each firm is producing at its capacity and there is no incentive for competition by undercutting its competitor. In case of excess capacity, mixed strategy equilibrium exists. This means that if the Nash game is repeated over time by the two firms, the model predicts price variations between periods.

Notice that in Varian(1980), some consumers are imperfectly informed and they only learn about prices over time and so price dispersion depends on imperfect information. Osborne and Pitchik(1986), however, attribute price dispersion to shape of demand; if demand is such that the profit function is concave, the equilibrium price varies in a certain interval; if the optimal inverse demand is outside the interval, the large firm either charges a high price or is likely to offer a substantial discount while the small firm tends to charge low price most of time. Thus in this case, the large firm randomly holds sales as in Varian (1980). However, the difference is that price randomization in Varian's model occurs when consumers are imperfectly informed of prices while that in Osborne and Pitchik's model emerges from the process of competition and not from imperfect information. Thus Varian's model explains price dispersion based on customer price information whereas Osborne and Pitchik's model illustrates the effect of competition on multiple pricing. Varian's model better explains price dispersion in credence services (e.g., medical services, legal services, consulting services, etc.) rather than Osborne and Pitchik's model since choice of credence services based on trust fosters long-term relationships that discourages active search in price, so imperfect information on price for credence service persists.

Implications of the Osborne and Pitchik's model can be explored by comparative statics. As the size of the large firm increases, its equilibrium profits are greater. On the other hand, as the size of the small firm increases, the greater competition may offset the direct effect on its profits from increased sales and thus the net effect is not determinate.

This point is consistent with Gelman and Salop (1983) which argues that even if capacity is free to install, a firm may not want to expand infinitely because the net effects of increase in capacity of a small firm are not clear; larger industry capacity is associated with low equilibrium prices while small size of a firm can reduce competition with large firms and lead to high equilibrium prices.

Service capacity means both physical capacity and time of consumption. For the same magnitude of physical capacity, service firms with greater customer turnovers would have larger effective capacity. If customers are heterogeneous in terms of consumption time, it pays to implement differential rationing strategies. von Ungern-Sternberg(1991) regards seating capacity with a view as a commodity, even if it is not charged, like drinks in an outdoor restaurant situation and explains why customers are frequently rationed at certain types of restaurants. Suppose there are two kinds of customers: business customers who are highly constrained by time and tourists whose consumption is quite time intensive. In the restaurant, business customers' demand do not take up much of its spare seating time and they are more price sensitive in this specific situation than tourists who take seats for a longer time and enjoy scenery. The restaurant should try to keep prices low on items ordered by the business customers and increase the prices of favorite items with time-intensive tourists for obtaining higher profits.

3.2.4 Service Reliability

In service businesses, when demand exceeds capacity of a firm, rationing of the excess demand often takes form of service denial or waiting line to be served. Uncertainty about whether or not customers are able to be served might affect utility of service and so service choice. Service reliability might therefore play a crucial role in service competition.

De Vany(1976) views waiting time as product quality. In his model, demand is random and production is also random. As a consequence, amount of customer waiting is uncertain. Even though the model has the perspective of a monopolist who is strictly preferred to a second best alternative producer by consumers, ceteris paribus, customer search is addressed such that if the waiting queue length exceeds a critical value, the customers do not join the queue and visit their next-best alternative firm. The critical queue length depends on a set of factors such as the distribution of queue lengths, cost or value of waiting time, cost of search and availability of alternative service firms. This customer rationality creates a search for optimal length of queues. The customers will tolerate longer queues, i.e., the optimal queue length is longer if price of the monopolist service is lower, price of the alternate is higher, service rate of the monopolist relative to that of the alternate is greater, cost of search for the next-best one (i.e., cost of switching to the alternate) is greater and cost of waiting time is lower. De Vany shows that change in price or capacity affects quantity and quality of service. In equilibrium, there is an excess capacity. This is quite compatible with observations in a real service industry because most service facilities except for a small number of very popular service firms operate below a full capacity even in peak periods. These results can be also found in Bulow et al.(1985) and Benoit and Krishna (1988) who argue that excess capacity in equilibrium acts as a threat to competitors. Even though price equals marginal costs, a monopolist would produce too little output and so carry too little capacity as long as she can enjoy positive profits. Although De Vany makes a contribution by considering customer waiting time, all consumers are assumed to have the same cost of waiting time. This assumption is too strong in the sense that cost or value of time might depend on factors such as one's income level and consequences of service purchase. One can expect that high-income customers have, on average, larger cost of waiting time than low-income customers. One may be also willing to wait more for medical service than for car repair service since the former has a more critical consequence than the latter. The cost of waiting time might also depend on the availability of alternative services, since for service with few or no alternatives in the neighborhood, the cost of waiting would be small or increase at a smaller rate as waiting time increases. For instance, we may observe this phenomena at the ticket office for special concerts or sports events like the Superbowl and the World Series because there is no alternative source of getting the tickets if resale between customers is ignored. Inclusion of service availability would lead to more effective pricing mechanism and capacity decision to reduce customer dissatisfaction from waiting. In sum, the above literature suggests that service availability should be incorporated into pricing strategy. Services that are readily available from a competing service provider should be priced competitively. On the contrary, services not readily available from other providers can be priced higher because customers should make further shopping efforts and pay costs needed to locate another firm for the desired service.

Within the symmetric Bertrand-Nash framework among customers, a representative firm and all other firms, Gerstner(1986) considers service reliability in the form of crowding costs. The crowding costs arise when customers visit a crowded firm and should wait in a long queue and be served at a slow rate. The costs are formulated as an increasing linear function of the number of customers at a service facility. These crowding costs are included in customer utility function and thereby affect service firm choice. The model also includes inconvenience cost of using the service at not most preferred time (i.e., at off-peak time). Even though Gerstner applies the peak load pricing to service industries, his major contribution can be to formulate explicitly the crowding costs and the inconvenience costs.

In the Bertrand Nash equilibrium, Gestner derives that a temporal price dispersion is optimal. The price difference between first off-peak and then second peak period is shown to depend on whether or not the firm commit itself to price over time. Under commitment, the firm makes pricing decision in advance and consumers are assured that prices cannot change later. Under no commitment, consumers know that price can change in second period. We can think of two cases: (1) customers have little or no preference for time of service use, e.g., service by hairdresser, medical service providers, etc.; (2) customers are reluctant to buy in the off-peak period and more customers visit firms in the peak period, e.g., service by hotels, airlines, restaurants, etc. In the first case and under commitment, customers would spread themselves equally over periods to minimize crowding costs and so firms have no incentive to offer discounts in the off-peak period. This leads to the results that peak price should be equal to off-peak price in equilibrium which means *uniform pricing* over time periods is optimal. In the first case and under no commitment, however, firms would be motivated to reduce price for competing better in the second peak period since the first off-peak period buyers have already purchased the service and cannot shift their demand to the second period. This can lead to lower price and higher demand in peak period that is often called *price reversal*.

3.3 Capacity Choice and Pricing in Service Competition

Service firms might decide on their capacities based on forecasts of future demand and competitors' capacities before they compete in price. This means that price competition given fixed capacity is a subgame ignoring capacity decision. It would be richer approach to take both capacity decision and price decision into account in competition. A model of price competition after capacity competition assumes that prices adjust more easily and quickly than capacity and that a firm's capacity is observed by its competitors before the pricing stage. If capacities are, however, imperfectly observable to competitors, the decision process is as if capacities and prices are chosen simultaneously, although they need not be in reality. Gertner (1985) analyzes simultaneous quantity-price setting games. Each firm i chooses a quantity q_i and a price p_i without first observing its competitor's choices. For simplicity, we can assume that there are two firms in the market and that producing q_i costs $c\ q_i$, i.e., constant return to scale. It follows that no pure strategy equilibrium exists. This is closely related with the Bertrand-Edgeworth spirit. Gertner shows that a unique mixed strategy equilibrium exists; with constant return to scale, firms make a zero expected profit while with decreasing return to scale firms make a positive expected profit. This is similar to Bertrand's equilibrium. It is also consistent with Cournot's equilibrium in the sense that the expected price exceeds the competitive price, which equals marginal cost. Unlike Cournot's equilibrium, however, if capacities are observable, a firm cannot choose a limited capacity not to flood the market. This causes competitive pressure and drives profits down as in the Bertrand equilibrium.

3.3.1 Basic Models

In this subsection, capacity as well as price is a decision variable to meet varying demand. When a firm offering lower price faces more demand than its capacity, the excess demand is rationed to firms with higher prices according to a rationing rule. Service reliability, customer search and switching costs will be discussed in next subsections.

Suppose two firms face a two-stage competitive situation. These two firms produce perfectly substitutable services. At the first stage, the firms simultaneously and independently for simultaneously build capacity for subsequent service operation. Then firm *i* would have the following profits depending on relative prices.

$$\begin{array}{lll} & \{\ p_i \min [\ k_i\ , D(p_i)\ ]\ \} - c(k_i) & \text{if } p_i < p_j \\ \Pi_i = & \{\ p_i \min [k_i\ , \max (\ 0, D(p_i) - k_j\ )\ ]\ \} - c(k_i) & \text{if } p_i > p_j \\ & \{\ p_i \min [k_i\ , \max (\ D(p_i)/2, D(p_i) - k_j\ ]\ \} - c(k_i) & \text{if } p_i = p_j \end{array}$$

where unit cost of production is assumed zero and $c(k_i)$ is unit cost of installing capacity.

Each firm solves unconstrained problem of expected profit maximization problem where at the first stage the firm choose capacity level based on predicted demand behavior and presumed but unobserved competitor's reaction, and then at the next stage each firm competes in price.

Levitan and Schubik (1978) model duopoly in which both price and quantity are strategic variables. They assume that demand is nonstochastic and aggregate demand comes from identical individual downward sloping demand curves and that consumers can move from one firm to another with no costs. In the model, when demand exceeds capacity of a firm, residual consumers are all rationed efficiently. Levitan and Schubik show that no pure strategy equilibrium exists; when the lowest price is above marginal cost, firms have an incentive to undercut and serve the whole market; if the price equals marginal cost, then firms would raise the price and reduce production. No equilibrium existence is due to the assumption that demand is not stochastic since if demand is deterministic, gains from deviating strategies by the firms in duopoly would be large. On the other hand, if we assume that demand is uncertain, equilibrium would exist as the number of firms increases infinitely since possible gain from deviating the equilibrium is small as in Peters (1984). In a services marketing context, if there is no uncertainty on demand and customers have no switching costs from current firm, firms would severely compete in price and charge prices close to marginal cost, so they just earn negligible positive profits or no profits. If demand is, however, uncertain or highly seasonal, firms could earn positive profits by deviating the equilibrium price.

Following Levitan and Schubik (1972), Kreps and Scheinkman(1983) assume that in a two-stage duopoly competition, the low price firm faced by excess demand rations its available capacity efficiently and thus always serve higher valuation customers ahead of lower valuation ones. At the first stage two identical producers decide independently and simultaneously how much they will produce. Then each producer learns how much the other firm will produce and the two firms engage in Bertrand-like price competition. Demand is then determined by the price competition and production takes place subject to capacity constraints generated by the first-stage decisions. Under the efficient rationing scheme of residual demand, they demonstrate that the unique perfect equilibrium outcome of the two stage competition in capacity and pricing is equivalent to the equilibrium of Cournot model where firms conjecture product quantities based on prior knowledge of the market structure and demand characteristics, and set prices so that the conjectured quantities are realized. This suggests that the Cournot equilibrium is equivalent to the result of price competition among firms if they choose the capacity of production prior to setting prices. That is, Cournot behavior can still persist in the two stage game and the Cournot model of competition in capacities can be interpreted as a kind of reduced form for capacity-constrained duopoly competition with endogenous price setting. In view of game theory, firms can coexist in a Cournot quantity-setting game as long as one firm cannot satisfy total demand. In a game of price competition, the competition will continue until a lowest price firm expels the competitors out of the market if the firm meet total demand; otherwise the firm would at least sell up to its capacity and a second lowest price firm would serve residual demand.

Capacity decision at the first stage is critical since Cournot behavior can still persist in equilibrium of a two stage game and optimally set prices in relation to the Cournot output do not significantly affect equilibrium profits. The suggestions may well defend that in Cournot's model prices are optimally set, which has been a rarely accepted concept since Bertrand's criticism of Cournot model. The rationing rule chosen by Kreps and Scheinkman is, however, not a robust one in the sense that the efficient rationing rule is impossible unless firms can observe different consumer valuations for the product. The presumed parallel shift rationing scheme also leads to the worst contingent demand, $D(p_2 | p_1)$, for high-price firm 2 partly because each customer's willingness to pay is decreased equally and partly because the low-price firm skims the most attractive group of customers with high valuations in the market. The findings of Kreps and Scheinkman can explain why in spite of severe competition, firms producing identical service coexist and earn positive profits in the market. For eliciting customer valuation of service, firms may take reservations from high valuation customers, as in PNG (1989), and serve them first in priority to obtain high profits.

Davidson and Deneckere(1986) point out that the efficient rationing used by Kreps and Scheinkman(1983) requires firms' information on customer valuation which is often private information to customers but not observable to firms. The model considers an alternative rationing rule of residual demand. When customers have identical downward-sloping demand curves, this rule rations output on a first-come first-served basis and allows customers to make unlimited purchases as long as output is available. Notice that in Kreps and Scheinkman, there is a limit on the number of units to purchase at low price firm so that each customer can buy a part of what they want. This may reflect more reality in service industries. The customers who leave for the high-price firm would then be a random sample of the customer population, and thus sales for the firm two would be $D(p_2 \mid p_1) = D(p_2)\ [1 - (k_1 / D(p_1))]$ for all $p_2 > p_1$. This specification is different from Kreps and Scheinkman (1983) in that the high price firms' ability to charge different prices increases as a more favorable contingent demand curve is left for the high-price firm. Then for given values of capacities, firms earn larger profits. The attractiveness of larger profits makes firms compete more intensely and results in higher equilibrium capacities. Under the more realistic rationing rule, Davidson and Deneckere show that unless capacity is very expensive, optimal output is higher and optimal price is lower so profits are substantially lower than corresponding Cournot equilibrium levels. This suggests that the two-stage equilibrium is more competitive than the one-stage Cournot outcome that contradicts the main results of Kreps and Scheinkman. The Cournot equilibrium then sets an upper bound on the degree of competitiveness. When capacity is relatively cheap, each firm chooses different capacity although the model is symmetric duopoly. This suggests that there might be a tendency toward capacity asymmetries in such markets. This also explains why firms usually have different amount of capacities in the service industries. Another interesting results of the model are that in equilibrium firms use mixed strategies in price setting. The mixed strategies can be interpreted from two perspectives in the literature. Noting that price dispersion is pervasive in most markets, Varian (1980) points out that any model which yields pure strategy seems to be rather implausible explanation of price dispersion since consumers learn about prices over time and stop visiting high-price firms. From Harsanyi (1973)'s perspective, on the other hand, the pure strategy is a limiting case of the mixed strategy equilibria as random disturbances about profit functions go to zero. Firms' profit functions are subject to random shock due to incomplete information on production costs. In a service market where firms do not usually take reservations, Davidson and Deneckere model is more realistic since firms in the market do not know distribution of customer valuation and serve customers on first-come first-serve rule. Service firms adopting random serving principle would price lower but produce larger than firms implementing reservations system and priority serving principle, according to the results of Davidson and Deneckere model.

Benoit and Krishna (1987) extended Kreps and Scheinkman(1983) in a dynamic framework. Their model involves competition among firms repeated over an infinite time horizon. In a infinite time horizon duopoly where firms' choice of initial capacity is irrevocable, they find that firms have excess capacity in equilibrium. This result is contrary to Cournot equilibrium. The excess capacity has an important effect on prices and profits.

Profit-maximizing optimal prices with excess capacity would be higher than competitive equilibrium price. The reason is that some customers with irregular demand should pay higher price since their demand necessitates idle capacity. In this way, profits with excess capacity would be higher than Cournot level. The equilibrium price with excess capacity would, however, be still lower than price charged by a monopolist. Firms cannot choose smaller capacities to avoid excess capacity since each of them has incentive to expand given that the other would not be able to retaliate. If both firms, however, choose larger capacities, it is unprofitable for them to use these fully. For maximizing profits, both firms would sell at a level below their capacities. The main role of excess capacity is to present a threat to a potential deviating firm. This argument is also found in a repeated game paradigm, (for example, Bulow, Geanakoplos and Klemperer 1985). In this pure repeated game, players incur no costs when they make credible threats; a cost would occur only if the threats had to be implemented and without exogenous shocks, this never happens in equilibrium. On the other hand, in Benoit and Krishna's model, since firms' choice of capacity at the beginning is not revocable, it is costly to make threats credible, i.e., cost of excess capacity, even if these threats never need to be carried out. When firms, however, have the ability to continuously adjust their capacity levels, however, they may be able to earn higher profits than Cournot level as long as discounting factor is not severe over time. Even a firm without any initial excess capacity can always expand later, thus providing a lagged potential threat to its competitor. However, when firms have limited flexibility after initial capacity choice, excess capacity still acts as a preemptive deterrent to a capacity expansion by deviant rivals. For summary, Benoit and Krishna argues that a firm with excess capacity can earn higher profits than Cournot level (level by Kreps and

Scheinkman model) and than level by Deneckere and Peck model, but lower profits than monopolist's level. This argument may be the case especially when there is a high demand uncertainty, e.g., seasonal services because firms are not flexible to adjust their capacities to fluctuating demand in the short run.

From hoteliers' perspectives, Sheel (1995) demonstrates a decision-making tool by using Monte Carlo simulations. A Monte Carlo simulation is a way to consider all feasible mix of decision variables and exogenous variables and to perform all the possible "what if" analyses. Hotel managers should often decide on room mix (i.e., the number of rooms for each type like economy, business, executive, etc.), total number of available rooms and room rates. Managers should also be aware of customers' price sensitivity for each room class; customers for economy class are usually more price-sensitive than for executive class. Price elasticities of demand for each room class are given exogenously so that price elasticities for economy, business and executive (in a descending order in absolute values) classes are negative and elasticity for gold-card is positive. Under three possible scenarios of best, moderate and worst case in terms of room revenues, Sheel makes scenario analysis of capacity conditions to find optimal total number of rooms and then analyze price mix and room mix conditions given the total room capacity. The results of the simulation and scenario analysis provide hotel managers with guidelines to determine an optimal price, room mix and room capacity, given a set of price sensitivities of demand. The guidelines may vary depending on room-revenue scenarios; optimistic (upper 5th percentile), moderate (middle 50th percentile) and pessimistic (lower 5th percentile). Final choice, however, would depend on managerial judgment, manager attitude toward risk; an aggressive management would choose to strive for the optimistic scenario with full notion of small chance (5 percent) of achieving it while a more conservative management would rather strive for the moderate scenario with broader 50th percentile. Usual caveats may apply to Sheel's simulations. When reality is beyond the range set by the analysis, the results of the simulations can be local optima and are not 100 percent accurate. Another drawback of Sheel's analysis is that the author does not consider costs of expanding total capacity and rather implicitly assume that the hypothetical hotelier has enough capital for capacity expansion. When capital investment for adding room capacity accounts for a major portion of total costs for hotel management, costs of expanding capacity may affect significantly price mix, total room capacity and room mix.

3.3.2 Service Reliability

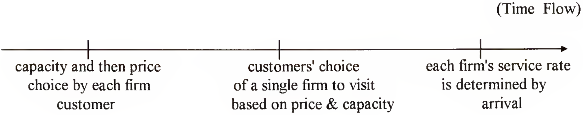
In this subsection, models take reliability of service into account in capacity and pricing decision. Services firms may consider reliability in making strategic decision in capacity and price since capacity affects probability of buying service at a specific time and pricing should depend on reliability as long as customers have different willingness to pay for different reliability.

From an agency-theoretic perspective, Png(1989) develops a model of pricing and capacity choice. Customers are risk averse and have private information on their valuation for use of capacity (there are two types of customer valuation: high and low) but only after service contract is signed. In the agency model, neither spot sales (purchase at time of use) nor advance sales (purchase in advance and no money back) will not maximize expected profits. Rather reservations system will maximize firm expected profits so that high valuation customers exercise reservation and low valuation customers or speculators do not. Thereby customers are induced to reveal their type of valuation. The reservations scheme provides high valuation customers with insurance against unavailability of capacity. This is similar to a screening game where firm provides a menu of contracts so that customers with private information self select an option. PNG model can apply to airlines and hotels that collect payment at the time of reservation and charge penalty for cancellation or pay compensation for no available capacity due to overbooking.

In a model of two-stage capacity choice and subsequent price competition, Herk (1993) analyzes firm behavior and consumer choice in a duopoly market for a homogeneous, nonstorable commodity, such as a service. Each consumer first attempts to purchase his entire demand quantity at the low-price firm. Then unsuccessful consumers who are not served try again at the high-price firm. Firms' residual demand is determined by firm-selection decisions of consumers who have rational expectation about the joint distribution of prices and service reliability across firms. Service reliability refers to probability of service outage or of customer being bumped from service. The model assumes that each firm chooses capacity in the first stage to maximize its equilibrium profits in the second stage net of capacity expenditure. Herk considers both of the rationing rules of residual demand: parallel shift residual demand specification analyzed by Kreps and Scheinkman(1983) and proportional residual demand specification examined by Davidson and Deneckere(1986). Nonetheless, he uses the former as a benchmark. Herk confirms the findings of Kreps and Scheinkman(1983) by deriving that if residual demand at the high-price firm less than or equal to that resulting from the parallel shift specification, or if marginal capacity cost is sufficiently high, the duopolistic capacity-constrained competition would lead to Cournot equilibrium by both firms in the second stage. On the other hand, if residual demand at the high-price firm is greater than that arising under the parallel shift specification and marginal capacity cost is relatively low, a mixed strategy pricing ensues in equilibrium. The latter case may explain why airline companies sell tickets at time of reservations and charge a nominal stand-by price for empty seats just before takeoff because marginal costs are so small, e.g., beverages and peanuts.

Capacity decision includes determining category and size. Planning capacity gets complicated under demand uncertainty. Paraskevopoulos et al.(1991) analyzes optimal capacity planning problem under demand uncertainty. The objective function is to minimize penalty of errors due to factors such as variations of demand forecasts, model misspecification, etc. Capacity expansion dealing with demand uncertainty can be addressed either on equivalent deterministic concept or stochastic capacity expansion models. To avoid the computational complexity of nonlinear stochastic control, the authors cast the capacity problem in a deterministic framework. It is shown that as reliability against demand uncertainty increases, the variance of profit function decreases. The authors view excess capacity in the face of uncertain demand as an outcome of an optimal investment strategy (see Lieberman 1989). In other words, optimal capacity expansion plans may involve some degree of excess capacity during some periods of increased uncertainty. As an empirical evidence, Overstreet (1993) reports 60 to 65 percent of US average occupancy rate for all kinds of hotels during 1982-1992 period, and points out oversupply and high competition in US hotel market.

In the market of n firms and a random number of consumers, Deneckere and Peck (1995) analyzes a two-stage game in which firms simultaneously select prices and capacities. They extend Peters (1984) by incorporating competition in service rate that results from the endogeneity of firms' capacities. Consumers know all prices and quantities but not the realization of aggregate demand. Their model is relevant in industries such as restaurants, toll roads and doctors' offices where facilities are congestible and demand is random. In their model, the timing of actions is the following:


(Time Flow)
capacity and then price choice by each firm customer
customers' choice of a single firm to visit based on price & capacity
each firm's service rate is determined by arrival

Each firm i simultaneously chooses its capacity, k_i, and the price of its output, p_i. After knowing price and capacity of each firm, each customer irrevocably chooses which firm to visit since it is very costly to visit a second firm. This situation might happen when transportation costs and costs of time for moving to another service firms are very high. Notice that unlike the other two stage competition models, firms choose their capacities before the customer arrival, i.e., without observing actual demand. At the time of customer arrival, each firm's service rate (i.e., capacity of firm i / total number of customers visiting firm i) is endogenously determined. If service rate is less than 1, some customers visiting the firm are rationed and the service rate indicates the probability of being served. In the model, demand is inelastic and so the firms cannot easily affect the way in which demand is rationed. In contrast with this assumption of demand, rationing rule influences

$$(\nu - p_i)\, s(k_i / q_i) = U^*(p, k) \qquad \text{for each } i \text{ such that } q_i > 0$$

$$\text{and} \quad (\nu - p_i) \leq U^*(p, k) \qquad \text{for each } i \text{ such that } q_i = 0 \text{ and } k_i > 0$$

where all customers have the same reservation value ν.

Given that q_i is implicitly determined via the above customer equilibrium strategy, firm i chooses an optimal strategy pair (p_i* , k_i*) to maximize profits.

Even though Carlton(1978) considers the same type of competition based on price and service rate, the contexts of Carlton is different from those of Deneckere and Peck. In Carlton's model firms are assumed to earn zero expected profits while in Deneckere and Peck's model firms strategically consider the effects of their actions on the market utility level and consumer behavior. As the number of firms increases infinitely, Nash equilibrium of Deneckere and Peck converges in the limit to the Carlton's results of competitive Nash equilibrium that industry profits approach zero. Their model also shows in the limit that the percentage markup over marginal cost equals the percentage of excess capacity. This is analogous to the general reusults of a profit maximization model that in equilibrium marginal revenue equals marginal cost.

Deneckere and Peck also examine the effects of changes in parameters on equilibrium prices and capacities.As marginal cost, c, decreases or customer valuation, v, increases, total capacity is higher and thus service rate is higher in equilibrium. The justification for capacity expansion might be that decrease in marginal cost and/or increase in valuation of service can allow firms to increase their profits by serving more customers and charging higher prices. Interestingly, increase in the number of firms, n, has no effect on the equilibrium total capacity and service rate, but decreases prices. This implies that when aggregate demand is less than total industry capacity and so there is a positive probability of unsold capacties, each firm has unsold capacities. Excess capacity in equilibrium will be discussed in a next subsection. On the other hand, when aggregate demand is greater than total capacity, each firm rations. More competition would then just lead to lower prices and thus lower profits. The effect of customer valuation on price is positive; higher ν allows firm to charge higher p for the same service rate.

The static solution of Deneckere and Peck's model has, however, some drawbacks in that consumers might learn firms' prices and service rates only gradually and firm's commitment to price and capacities can be a gradual process so that firms may have opportunities to react to one another. The Nash framework also lacks reality in service competition because each firm may have different capacity. This means it is plausible that some firms have to ration excess demand while at the same period other firms carry unused capacities. For instance, some popular restaurants face, on Friday night, high demand over seating capacity and ration excess demand by having customers wait in a long line or by bumping them; at the same time, other restaurants may have empty seats. It would be worthwhile to consider a dynamic paradigm, which includes customer learning and firm reactions, and/or solution concepts other than Nash.

3.3.3 Switching Costs

Service firms may set up a device, called switching costs, to discourage customers from searching for and switching to other firms. Switching costs can be *exogenous* or *endogenous* depending on whether they are intrinsic to the nature of the good or not (Herk 1993). For many services that are customized to buyers (e.g., medical service, legal services, financial services like investment banking, consulting services like accounting, diet programs, exercise clubs, etc.), there might exist exogenous switching costs such as learning costs in terms of time and energy taken for learning how to use a new service. Buyer-seller relationships in these industries are founded on intangible qualities such as trust and confidence. These kinds of services are often termed as credence goods (Darby and Karni 1973). On the other hand, even for highly standardized services (e.g., airline services, car rental services, fast food services, etc.), customer switching costs are created endogenously by firms' marketing strategies. These endogenous switching costs are designed to promote long-term relationships through buyer-seller contracts that penalizes switching or reward loyalty; for instance, airline frequent flyer programs, season tickets, book or music clubs and magazine subscriptions (Fisher 1991). For endogenous switching costs, posted prices by firms might usually be *contingent* on realized service availability so that the prices are only paid when consumers actually obtain service. However, it is also possible that once a customer signs a certain service contract with a firm, the price should be paid irrespective of service reliability; for instance, the customer pays *noncontingent* subsription price or access fee under a system of two part pricing to use cellular phone service.

Fisher (1991) interprets buyer switching costs as one of customer-based isolating mechanisms that provide firms with sustainable competitive advantage by building economic and psychological barriers to customer changing one firm to another. Fisher suggests that buyer switching costs are more effective isolating tools to firms seeking to maintain their current market positions since switching costs can provide the time to required to counter competitive strategies.

Considering cost-based brand loyalty such that the brand last purchased has an advantage or benefits, Wernerfelt (1991) derives single price market equilibrium. In the Nash symmetric equilibrium, price and profits are higher for slower search and lower market growth. Once service firms build up loyal customer base by creating switching costs, the firms are able to charge a high price to the loyal clientele and do not have to give multiple prices.

Herk (1993) analyzes the effect of customer switching costs strategically designed by a firm, on the equilibrium for capacity-constrained duopoly competition. In the model, each customer selects a firm that maximizes his utility with respect to price and service reliability. Once he has committed himself to a firm, the customer faces a cost of switching to another firm. Since switching costs are assumed to be sufficiently high, each customer makes a once-and-for-all firm selection according to his preferences over combination of price and reliability. Customers at the low price firm sacrifice service reliability for low price while customers at the high price firm are compensated by higher reliability. The firm selection decision by customers is consistent with firms' residual demand allocation when there is an excess demand over capacity for a low price firm. The low price firm attracts excess demand over its capacity until decrease in service reliability just offsets the price advantage, compared to the high price firm. In case where firms' posted prices are contingent so that the price is only paid when the service is available, Herk finds that in customers with switching costs and reliability preference ranging from risk neutrality to moderate risk aversion, the joint Cournot behavior provides the unique perfect equilibrium for capacity-constrained duopoly competition. Customer attitude toward risk here is related to demand specification for high price firm. That is, when buyers are risk nuetral or moderately risk averse with respect to service reliability, they are more willing to buy in the low price firm at the expense of reliability and thus residual demand at the high price firm would be less than or equal to that arising from the parallel shift specification by Kreps and Scheinkman (1983).

The above subgame perfect equilibrium involving Cournot behavior by both firms is in contrast with Davidson and Deneckere (1986)'s results that under contingent pricing, capacity-constrained duopoly with first-come first served rationing and no customer switching costs, exhibits higher aggregate capacity and lower expected profit than the Cournot solution. In other words, the capacity-constrained competition without switching costs would lead to more competitive equilibrium than Cournot equilibrium. These two contrasting outcomes imply that subscription pricing may facilitate tacit collusion in the capacity-constrained competition and that the upshot is lower total capacity and higher profits than those without any switching costs; this is analogous to frequent flyer programs in the airline industry (Banerjee and Summers 1987). In a real market of cellular phone services, a handful of major players in the market charges collusive rates and reap high profits via switching costs.

On the other hand, if buyers are highly risk averse with respect to service availability, in spite of switching costs customers would be more willing to search for another firm and thus more competition and higher demand uncertainty in the market would lead to higher total capacity for firms than the Cournot equilibrium. In reality, if one is very concerned about service reliability (e.g., he is inviting his financee to a Friday night dinner at restaurant), he is willing to search for all restaurants to get guarantee for service. Some prestigious restaurants may target these kinds of customers who are reliability-sensitive but price insensitive, and carry high capacity and/or take reservations

For the purpose of comparison, it is useful to review some switching costs literature without considering explicitly capacity constraints. The key issue here is the impact of exogenous switching costs on the degree of competitiveness of oligopolistic markets. Using Hotelling's model, von Weizsacker (1984) shows that if consumers are uncertain about their future preferences and if firms can commit to charging the same price in the following periods, the presence of exogenous switching costs makes markets more competitive. The reason is that uncertainty about future preferences together with switching costs increase customers' price sensitivity, which motivates firms to cut prices for increasing profits.

In a two period model, Klemperer (1987 a, b) shows that the results of von Weizsacker depends on the ability of firms to commit to a future price. Once one has been a customer of a particular firm, the customer may not be sensitive to competitive firms' price cuts as long as switching is high enough. Thus switching costs provide firms with incentives to start with very attractive prices for increasing customer base and to raise prices in the following periods, i.e., penetration pricing. If customers, however, have rational expectation that future prices charged by a firm increase with the level of its current market share, then they become less price sensitive to price cut in the first period. In this case, the presence of switching costs may even increase the level of prices in the first price.

Rather than precommitting to a price, firms can offer discount coupons which may specify absolute value or percentage terms. Again, switching costs are endogenized. Banerjee and Summers (1987) develop a two period duopoly model with homogeneous good where firms simultaneously announce coupons in the first period; i.e., discounts are granted to the first period consumers if they repeat purchases in period two. In their model, coupons act as a collusive device. In other words, without coupons, firms would earn zero profits in both periods because of Bertrand competition, but loyalty-inducing arrangements allow both firms to achieve a fully collusive outcome, i.e., joint monopoly prices in the second period by locking in customers with switching costs. As one of popular loyalty-inducing arranagements, frequent flyer program creates artificial switching costs by offering awards based on flying mileage and help form loyal customer base. Thereby airline companies can earn positive profits in spite of fierce price competition. The frequent flyer scheme might also deter new entry because new entrants do not have a frequent flyer base and so it is very hard and costly for new comers to attract customers who are loyal to incumbents.

Caminal and Matutes (1990) examine the issue of whether a specific form of firm commitment inflences differentially the pricing behavior. They show, in a two period duopoly model, that alternative forms of commitment result in qualitatively different equilibrium price dynamics and overall profitability. Price commitment involves a complete segmentation of the market in period two depending on consumers's choice in period two; loyal customers and newcomers. The price that a firm charges to newcomers does not have an impact on the second period profits earned form loyal customers since firms can charge a higher price to loyal customers with higher willingness to pay. This implies that each firm has an incentive to be very aggressive in period two if the other rival firm precommits to a low price. This is not the case with coupons since coupons does not allow the firms to completely segment the market in period two as price commitment does. That is, the price chosen in period two clearly has an impact on the profits earned from loyal customers since loyal customers have also chance to use the coupons. In short, price commitment implies lower profits for both firms while coupon commitment involves higher profits. Thus price commitment tend to enhance competition while coupons tend to decrease the competitiveness of the market.This is consistent with Banerjee and Summers (1987) which find collusive prices between airline firms given that firms are allowed to offer free tickets or upgrade coupons in frequent flyer programs.

3.4 Implications from the Literature

3.4.1 Price Dispersion: Good or Bad

The classical law of one price posits that frictionless competition will force identical items to be sold at the same price; however, in fact homogeneous goods often sell for widely varying prices (Dahlby and West 1986; Carlson and Pescatrice 1980; Pratt et al. 1979; Stigler 1961). Is price dispersion good or bad for customers or for service firms? The answer to the question might depend on the factor that requires multiple prices to exist in equilibrium.

When customers are not homogeneous in terms of preference for a combination of price and service reliability. Some consumers prefer to be relatively certain that they will obtain the service at the desired time in a particular place without much waiting and/or shopping. Such customers will choose to patronize higher price firms which offer a high probability that the service will be available. In this segment of market, the ratio of number of customers to number of firms is low. Other customers may prefer a relatively low price although they are aware the fact that there is a great chance that the service is not available. These customers will choose to patronize lower price firm with lower service reliability. In this segment of market, there is a relatively high density of customers per firm. If service is available, customers visiting a high price firm should wait more than those visiting a low price firm until a full price, posted price plus cost of *expected* waiting time, does not exceed a threshold, for example, reservation price. However, if service is not available or the full price exceeds the threshold, the customers who have visited the low price firm would leave for another firm. Searching for another firm after visiting a firm, however, would also takes transportation cost and time. Customers need to consider the trade-off between cost of further search and cost of waiting time. Thus customer preference for service reliability is closely related to cost of waiting time and search costs associated with transportation costs. Price dispersion based on customer heterogeneous preference for a combination of price and service can be desirable to customers since it allow each customer to *self select* a utility-maximizing combination out of a menu of price and service reliability combinations.

Demand uncertainty imposes additional costs on firms; when demand is not certain, a firm is not able to predict exact demand in a certain period and thus it should have potential idle capacity as a buffer. As usual, suppose that average total cost is U-shaped (i.e., average cost decreases up to a certain point of production and then increases after that point due to need for extra capacity, material, personnel, etc.). When demand is uncertain, firms are operating in the region of increasing returns to scale in the stochastic equilibrium because firms are not able to determine an efficient production, (i.e., decreasing average cost) based on uncertain demand. Thus expected profits could increase if forms can forecast demand with certainty. Both cost effect and profit effect provides firms with incentives to stabilize random demand by increasing the certainty of demand. For these reasons, firms offers lower prices to loyal customers and regular customers. This can explain why newspaper companies and megazine companies usually offer deep discounts to subscribers. Another example is that airline companies also offer very attractive prices to the customers who buy the tickets in advance of departure date.

However, other customers whose service purchase patterns are hard to predict, e.g., business travellers and irregular readers, should pay higher prices because their irregular demand causes demand uncertainty for the firms.

Buyers may have no perfect information on price distribution of a certian service in the market. The price distribution of the same service can be created by the heterogeneity of the firms' operating costs plus capacity costs. Firms have the private information on cost. There exists information asymmetry in price between firms and customers. Some customers who has imperfect information on price distribution would pay higher prices than others with perfect information. Customers' search costs can also turn each firm into a local monopolist. When customers do not make a complete search due to high enough search costs, they might not be able to obtain the best possible price offer. The price dispersion based on customers' imperfect information on price distribution may not be desirable because firms price discriminate customers only for heterogeneity of price information. A reason for cutomer imperfect information migh be inertial brand loyalty where there is no effect of brand last purchased on current choice and consumers are slow to become aware of brand with highest value (Wernerfelt 1991). Inertial loyalty is in contrast with switching costs-based loyalty that is highly dependent on previous purchase. Examples of inertial loyalty can be observed in purchase of low-involvement services that do not require active search for alternatives in a class. For instance, people do not search much about for lunch.

3.4.2 Excess Capacity in Equilibrium

From a strategic point of view, firms may build capacity in excess of what is needed along the equilibrium path to deter new entry into the market or to pre-empt existing competitors (Lieberman 1989, 1987; Benoit and Krishna 1987; Eaton and Lypsey 1977). Excess capacity enables incumbent firms to threaten to expand service output and cut prices following entry so that entry appears unprofitable. In this case, existing firms use excess capacity as a credible signal to deter entry. This is analogous to limit pricing (Milgrom and Roberts 1982) where incumbent firms signal their private information about costs (marginal costs associated with production or fixed costs with capacity). Excess capacity also acts as a threat to existing competitors who may try to deviate from the equilibrium price and output for higher profits. This is analogous to a war of attrition that reflects battle for survival among firms currently active in a market (Nalebuff and Riley 1985). On the other hand, when the market is expanding, a firm with excess capacity has advantage over the other firms without excess capacities because firms have limited flexibility to adjust capacity to demand. It is, however, costly to making threats credible by making investment in excess capacity prior to new entry or competitive deviation. Thus to determine the degree of excess capacity, firms should consider trade-off between creating credible threat and making substantial investment.

Service firms also need to have excess capacity when demand is uncertain (see Lieberman 1989; Paraskevopoulos et al. 1991). Demand uncertainty makes it difficult for firms to forecast demand in a certain time period. Excess demand over capacity in a certain period would be bumped out of service, thereby reducing possible firms' revenues. If firms try to avoid possible service outage, firms would rather have excess capacity as a buffer. This is the case in particular when service denial due to outage lead to a permanent loss of business. Thus in the presence of dema2nd uncertainty, firms would have more capacity than expected demand. This leads to excess capacity in equilibrium. Therefore in the face of demand uncertainty, excess demand may not be evidence of inefficient investment but an outcome of rational investment as an optimal response to uncertainty.

3.5 Conclusions and Future Research Directions

In the conceptual paper, we have reviewed research both on competition in price with capacity constraint and on competition in capacity and price. In the short run, firms can only adjust prices to demand and they compete in price given fixed capacities. In the long run, however, firms have flexibility to adjust capacity and price to demand, and they compete both in capacity and price. The research of which we make review is very relevant to services competition since non-storability of service makes capacities be a primary strategic consideration to services firms. The relevant literature has been analyzed on the dimensions of service reliability, customer search and switching costs. Service reliability is closely related to firms' decision on capacity and operation; thus this dimension is supply-side. Prior search that customers make before visiting a firm, is closely related to search costs that are customer specific. Posterior search after the visit is related to cost of expected waiting time at the first firm and of further search costs such as transportation costs and time; thus this dimension is demand-side. Switching costs can be exogenous to a certain service; this is customer specific like learning costs. Switching costs can be also strategically endogenized by service firms that seek to increase loyal customer base; this is initiated by a firm, e.g., frequent flyer program, subscription pricing, etc. Hence the dimension of switching costs is a mix of demand-side and supply-side.

Table 3-2 Classification of the Core Literature on Capacity and Pricing Decision

	Price Decision given Capacity (Short-run)	Capacity and Price Decision (Long-run)
Basic Models	Osborne and Pitchik(1986) von Ungern-Sternberg(1991)	Levitan and Schubik(1972) Kreps and Scheinkman(1983) Davidson and Deneckere(1986) Benoit and Krishna(1987) Sheel(1995)
Service Reliability (Supply-side)	Carlton(1978) Gould(1978) Davidson(1988) Arnold et al.(1989) Jones and Hamilton(1992)	PNG(1989) Paraskevopoulos et al.(1991) Herk(1993) Deneckere and Peck(1995)
Customer Search (Demand-side)	Peters(1984) Bryant(1990)	
Switching Costs (Mix of Demand & Supply-side)		Fisher(1991) Wernerfelt(1991) Herk(1993)

The key relevant literature can be arranged in the following Table 3-2. From the table, we notice that most research focuses on service reliability. Little research has been done in the area of customer search and switching costs. Since customer search for service firms based on price and service reliability includes search costs and time, inclusion of search behavior in modeling services competition leads to different equilibrium price, capacity and profits from those in models ignoring search.

Switching costs that can be exogenous and /or endogenous are also very interesting. More specifically, non-contingent (on service availability) endogenous switching costs can justify the Cournot subgame perfect equilibrium for two stage game of capacity and price competition; this suggests that firms could have positive profits in equilibrium since they compete in service output. Switching costs are closely related to firms' loyalty programs that keep on being sophisticated for expanding loyal clientele. Varying loyalty programs can influence differentially switching costs and thus have a differential impact on firms' pricing and capacity strategies. Customer might also consider switching costs for search and final firm choice. Therefore consideration of both search and switching costs, and their interactions would make a model on service competition richer.

One implication that can be drawn from the literature is that there exist multiple prices for the same kind of service in equilibrium. Out of the factors that make multiple prices exist for the same service, price dispersion is good for customers when they are heterogeneous in preferences on a combination of price and service reliability (or service availability.) The reason is that customers are allowed to self-select a combination from the menu so that their utilities are maximized. Price dispersion related to different tastes for a set of price and reliability is unique to services since services cannot be inventoried. Thus modeling this kind of multiple price equilibria in services market would provide an interesting insight to services managers. Demand uncertainty also can lead to price dispersion. In the presence of demand uncertainty, for avoiding service outage, firms should have excess demand that creates additional costs and thus damage profitability. Loyal customers who purchase service regularly should be allowed to pay lower prices than irregular buyers. Likewise those who buy the service in advance of actual time of service can allow firms to reduce excess capacity should be able to pay lower prices than those who do not. When consumers do not have perfect information on price distribution for a service in the market, firms may be able to price discriminate customers who have varying degree of imperfect information. High search costs and thereby limited customer search may also turn firms into local monopolists who are able to price discriminate. In this case, price dispersion is not advantageous to customers since the "monopolists" would be able to extract consumer surplus.

Another implication from the literature is excess capacity in equilibrium. Excess capacity may deter new entry into the market and also preempt incumbent competitors trying to deviate from the equilibrium prices; i.e., excess capacity acts as a strategic threat to possible new entrants and competitive incumbents. Demand uncertainty also pose need for excess capacity to firms in particular when service denial due to lack of enough capacity could have a substantially detrimental impact on firms' profitability; e.g., permanent loss of business.

An interesting direction for future research is to take peak load pricing into account in the area of services competition. Peak load pricing is usually used for shifting peak demand to off-peak period by firms that intend to make a more efficient use of capacities. When demand is fluctuating in a predictable pattern (i.e., peak period and off-peak period), service firms do not only compete in the same period but also in a different period. The reason is that peak demand is highly related to off-peak demand both within firms and across firms. That is, in some sense, a firm's service at peak is a substitute for the same firm's service at off-peak and also other competitors' service at peak and off-peak. Inclusion of peak load pricing into service competition would also influence firms' equilibrium strategies in price and capacity. When we include peak load pricing into service competition, there may be two cases. For simplicity, suppose there are two players in a service market, i.e., duopoly. One is a scenario where given capacity constraints, a firm adopting peak load pricing competes with its competitor charging a single price for the service in two periods of peak and off-peak. This scenario may give insights on under what conditions about demand characteristics (e.g., price sensitivities, seasonal variations) a service firm should adopt peak load pricing or use uniform pricing. Another scenario is that two firms adopting peak load pricing compete with each other given capacity constraint for each firm. This scenario may provide managerial implications on how service manager choose peak and off-peak price in relation to competitor's peak and off-peak price under a certain demand and capacity condition.

Reservations system is also a promising research issue in services competition. Since reservations policy by a firm is closely related to its service reliability. Customers who are risk averse with respect to service availability at a desired time would be more willing to make reservation, even though they may pay at the time of reservation, prior to actual service, and may face cancellation penalty in case that they cannot use the service at the specific time. Risk neutral customers then may go without reservations. Thus the trade off between service guarantee and cancellation penalty that reservations system can provide, may affect differentially customer behavior of making reservations depending on customers' attitude toward risk with respect to reliability. In other words, service firms can implement reservations system as a device of allowing for customer self-selection. Reservations system is popular in some areas such as airlines, hotels and car rental companies but is not usual in other areas such as fast-food restaurants, bars and emergency care centers. Thus it would be very interesting to investigate the following questions; what type of service industry with a certain demand pattern should adopt reservations systems. This question could be examined in a screening game situation: customers have private information on time preference and its valuation, so a service firm will offer a menu of combinations of price and service time and also reservation options. Another related research questions are on which firm in terms of capacity size and cost structure within a certain service industry should implement reservations policy; if a firm with a certain capacity and cost structure should use the system. Another related question is how much cancellation penalty the firm should impose and to what extent of overbooking the firm should take. These questions could be investigated in a maximization problem, i.e., how a service firm that has certain capacity and marginal costs as a percentage of capacity depreciation costs implement reservations system to maximize or allocate limited capacity to reservation holders to maximize its profits.

CHAPTER 4

COMPETITIVE CAPACITY DECISION AND PRICING IN SERVICES

4.1 Introduction

Demand for service fluctuates over time as demand for manufactured goods does. Besides random variations, service demand has predictable changes such as high season and low season demand. Since services are, in nature, not storable for sale in next period, unused or unsold service capacities have no value for firms at all (Sasser 1976; Davidson 1988; Png 1989; Deneckere and Peck 1995). For instance, this phenomenon can be observed in empty seats in a flight, unoccupied hotel rooms, non-rented cars and so on. On the other hand, when demand exceeds a firm's capacity, the firm would lose excess demand to its competitors.

Capacity is important in particular when firms should invest a large amount of capital in building capacities to serve visiting customers and thus they have limited flexibility in the short run to adjust capacities to demand, as compared with other decision variables. For instance, in the lodging industry, hotel construction boomed in 1996, but developers are cautious in continuing the trend for more building in the latter 1990s because the proper market and hotel category must be considered to avoid experiencing problems associated with overbuilding in the1980s (Wolff 1997). Overbuilding may lead to an inefficient capacity use and thus the profitability of service firms would substantially decrease.

Some literature in marketing and economics examines services pricing. Pricing bundle of services is investigated (Venkatesh and Mahajan 1993), and pricing services with demand seasonality (Radas and Shugan 1996; Gerstner 1986) and pricing services with random demand (Bryant 1990) are also analyzed. These researches do not explicitly take into account service capacity decisions. Capacity decision has been examined in operations research as the newsboy problem in which a firm chooses an inventory or production level for perishable good. The newsboy problem is investigated in a competitive setting (Lippman and McCardle 1997) and it is also applied to order quantity problem for value-decreasing goods such as style good in the retailing industry (Mantrala and Raman 1996). However, none of the existing literature has investigated the interactions of capacity, pricing, competition, demand seasonality and random demand fluctuation in an empirical setting.

The objective of the paper is to provide insights into capacity decision and pricing in the services industries where demand varies seasonally and randomly. That is, the proposed model has as its purpose of clarifying what are important factors in capacity and pricing decision and how the relevant factors affect the two decisions under certain conditions. In the subsection two of this paper, we propose a comprehensive model of optimal capacity and pricing in case where service demand varies seasonally. In subsection three, we examine the effects of competition and price sensitivities on service capacity choice. In subsection four, we investigate the effects of competition, capacity level and price sensitivities on service pricing decisions. In subsection five, we extend the model into a stochastic demand case and examine the effect of demand fluctuations on capacity decision. In subsection six, for validating analytical results of the model in the real services setting, we collect the data in the Florida hotel industry and do empirical analysis. In the final section, we summarize the findings and suggest the directions of future research.

4.2 Model Formulation

Although service demand can vary randomly, we begin with deterministic demand as a first level of analysis to gain some understanding the issues of interest without encountering complications. We tackle a comprehensive model of capacity decision and pricing in a situation where service demand varies predictably (i.e., high in season and low in off-season). There might exist multiple periods in terms of magnitude of demand. For instance, at the area of the Daytona Beach, Florida, there are multiple periods such as winter/summer low (Jan.5 - Jan.31 and Aug.20-Dec.15), winter high (Feb. 1 - Apr. 30) and summer high (May 1- Aug.19). Hotel and motel rooms are priced in the order of summer high (highest), winter high, and winter/summer low (lowest) (Delta Airlines Brochure, 1996). However, for simplicity, the model considers two periods: season and off-season. When season demand hit the capacity, excess demand may be denied for service. For instance, tickets for a big event such as the Super Bowl may be sold out so that some people might not be able to get tickets unless they pay higher prices to resellers.

For identifying relevant factors to capacity and pricing decision in the services industries, we investigate a duopolistic case where two services firms compete with differentiated services, to some extent. In general, a service is hard to standardize and there might exist many variations of one service. For instance, at the same level of overall hotel quality[1], each hotel may have different kinds of amenities such as cocktail lounge, heated pool, saunas, exercise rooms, etc. Thus the two services are not perfect substitutes. In sum, two players in the service market play a Nash duopoly game with differentiated services. We assume that service purchase is made once in season or off-season period, but not both periods. In other words, customers make choice decision on what service is purchased when, by comparing prices of competitive services between periods. We also assume that in season period, demand hits service capacity while demand is a way below the capacity in off-season period. For example, customers fill up the capacities of hotels in the Disney World Resort during holiday seasons such as Christmas or New Year's Day. On the other hand, in low season, hotels have low demand and thus some empty rooms.

Notation

i : firm i, i=1,2

a_h: location parameter for high-season demand

a_l : location parameter for low-season demand

b_h, b_l : price sensitivity of customers in season and off-season, respectively

γ : effect of competitive price on demand in the same period

θ : effect of own price on demand in cross periods

p_{hi}, p_{li} : season and off-season price of firm i, i=1,2

K_i : capacity of firm i

c : unit operating costs

k : unit capacity costs.

---

[1] According to the lodging evaluation criteria by the American Automobile Association, ratings are assigned, depending on the property's classification. The classification represents the physical design and

Based on the above notation, high-season demand and low-season demand for firm *i* are[2]

High-Season Demand $D_{hi} = a_h - b_h p_{hi} + \gamma\,(p_{hj} - p_{hi}) + \theta\,(p_{li} - p_{hi})$ (4.1)

Low-Season Demand $D_{li} = a_l - b_l p_{li} + \gamma\,(p_{lj} - p_{li}) + \theta\,(p_{hi} - p_{li})$ (4.2)

All the parameters are set to be positive. a_s and a_o represent the locations of high-season and low-season demand so that demand may shift in or out according to decrease or increase in the location parameters, irregardless of price change. Demand shift can be due to change in customer economic conditions such as income and employment as well as change in customer taste. In each period, firms 1 and 2 compete with each other and the cross-firm price effect is captured by γ. Given the assumption that purchase is made once, service in one period can be a substitute for service in the other period and the cross-period price effect is captured by θ. Since the services are assumed not to be perfect substitutes, cross price effect is smaller than own price effect (i.e., $b_o > \gamma$).

A two-stage model is investigated where service firms choose the capacity level in the first stage and then in the second stage, given the chosen capacity level, they decide on

---

level of services provided by the property; for example, a motel offers limited services while a resort hotel offers extensive guest services and recreational facilities (AAA Tour Book 1997).

[2] The demand function used here is an extension of Coulan and Mantrala (1992) by considering two periods. This type of demand function does not suffer from the problem of increase in total market demand with respect to increase in cross-firm (γ) and cross-period (θ) price sensitivity. An alternative specification, $D_{hi} = a_h - b_h p_{hi} + \gamma\, p_{hj} + \theta\, p_{li}$, may have the above problem as in McGuire and Staelin (1983).

price to maximize their profits[3]. In the two-stage decision with output choice first, an equilibrium does exist and output has a primary strategic importance (Friedman 1988). In the service industries, this sequential model with capacity decision first is realistic since service firms may build capacities in the beginning and keep them in the short run, and then the firms may be able to adjust prices for matching the capacities to demand. In contrast, a two-stage model with price choice first and then capacity choice might be more relevant in a retailing situation where order quantity by a retailer is flexible. The two-stage sequential model is also supported by the literature of game-theoretic models (see Kreps and Scheinkman 1983; Davidson and Deneckere 1986; Herk 1993).

We investigate relevant factors that affect optimal capacities and prices. In the first stage of proposed model, each firm decides on its service capacity. Then with capacity levels known to public, firms choose prices competitively. The way of solving the sequential model is via backward induction. That is, each player does pricing subgame given a level of capacity (i.e., optimal prices are functions of the capacity level) and goes back to the first stage of determining the optimal capacity. The rationale for the backward solution is that the equilibrium of the whole game should hold in every subgame, i.e., subgame perfect equilibrium.

The profit functions for firm i are

$$\Pi_i = \min[K_i, D_i](p_{hi}-c) + \{a_l - b_l p_{li} + \gamma(p_{lj} - p_{li}) + \theta(p_{hi} - p_{li})\}(p_{li}-c) - k K_i, \; i \neq j \qquad (4.3)$$

---

[3] A two-stage decision with capacity choice first is supported in Friedman (1988). In the paper, there is no pure-strategy equilibrium in a simultaneous game of price and output choice. For a two-stage game, either with output choice first or with price choice first, equilibrium does exist

The term in the bracket denotes the sales in high-season period. When high-season demand exceeds the capacity, sales are equal to capacity level. When high-season demand is below the capacity, the demand is realized as high-season sales.

Depending on relative magnitude of demand and capacity, we may suppose two cases. Case one is that both high-season demand and low-season demand are below capacity level. Case two is that high-season demand hits capacity while low-season demand is less than capacity level. In the former case, since even season demand does not hit capacity, capacity does not affect prices. In this case, the profit functions are

$$\Pi_i = D_i (p_{hi}\text{-c}) + \{a_l - b_l p_{li} + \gamma (p_{lj} - p_{li}) + \theta (p_{hi} - p_{li})\} (p_{li}\text{-c}) - k K_i , i \neq j \qquad (4.4)$$

The optimization problem is reduced to an unconstrained profit maximization one. This game boils down to Bertrand pricing game with no limit of quantity. Technically, optimal prices in the second stage are not functions of capacity levels.

Since, however, both capacity decision and pricing are of interest in the model, we only deal with the second case. Thus we consider the second phase of the two-stage Nash duopoly model, two service firms set optimal high-season and low-season prices given any capacity level. For the sake of a simpler calculation, we assume a symmetric equilibrium in which two firms have the same scale of capacity and charge the same price. Then the profit function would be

$$\Pi_i = K_i (p_{hi}\text{-c}) + \{a_l - b_l p_{li} + \gamma (p_{lj} - p_{li}) + \theta (p_{hi} - p_{li})\} (p_{li}\text{-c}) - k K_i , i \neq j \qquad (4.5)$$

Differentiating each firm's profit function (4.6) with its own high-season price, we solve the two first order conditions simultaneously to get symmetric high-season price for the duopolists.

$$p_{hl} = \frac{\left[K + (b_h + \gamma)(c + k)\right](2b_l + 2\theta + \gamma) + \left[a_l + (b_l + \gamma)c - k\theta\right]\theta}{(b_h + \gamma + \theta)(2b_l + 2\theta + \gamma) - 2\theta^2} \tag{4.6}$$

From the second order condition, all first-order minors of the Hessian are negative ($b_s b_o - \theta^2 > 0$) and all second-order minors of the Hessian are positive, i.e., $(2b_l + 2\theta + \gamma)(b_h + \theta + \gamma) - 2\theta^2 > 0$. Plugging p_{ol} into the first order conditions, we get

$$p_{ll} = \frac{2\theta\left[K + (b_h + \gamma)(c + k)\right] + (b_h + \gamma + \theta)\left[a_l + (b_l + \gamma)c - k\theta\right]}{(b_h + \gamma + \theta)(2b_l + 2\theta + \gamma) - 2\theta^2} \tag{4.7}$$

Moving backwards to the first stage of capacity decision, we substitute the high-season and low-season prices from the pricing subgame into the original profit maximization problem and solve for the optimal capacity.

$$K_l = \frac{a_h\left[(2b_l + 2\theta + \gamma)(b_h + \gamma + \theta) - 2\theta^2\right] + \theta(b_h + \gamma + \theta)\left[a_l + c(b_h + \gamma) - k\theta\right]}{(b_h + \gamma + \theta)(2b_l + 2\theta + \gamma) - 2\theta^2 - 2\theta^3}$$

$$+ \frac{2\theta^3\left[(b_h + \gamma)(c + k)\right] - \left[(b_h + \theta)(b_h + \gamma)(c + k)(2b_l + 2\theta + \gamma)\right] - \theta\left[a_l + (b_l + \gamma)c - k\theta\right]}{(b_h + \gamma + \theta)(2b_l + 2\theta + \gamma) - 2\theta^2 - 2\theta^3} \tag{4.8}$$

4.3 Services Capacity Decisions

Capacity costs affect high-season price and low-season price. Thus capacity costs are included in both prices. Despite the fact that low-season demand is below the capacity, since demands are interdependent between the two seasons, capacity costs affect both high-season and low-season prices. This result is in contrast with the general findings of the peak load pricing literature (Platt 1981): i.e., capacity costs are only reflected on peak price since capacity is determined by peak demand. The discrepancy may arise from the difference in assumption on demand that high-season demand and low-season demand are interdependent in the proposed model while they are independent in the general peak load pricing models.

Optimal capacity is a positive function of magnitude of both high-season and low-season market size ($\partial K_t / \partial a_h > 0$, $\partial K_t / \partial a_l > 0$ in Eq.(4.8)). Since high-season demand is inter-related with low-season demand through cross-period price sensitivity, capacity decision should depend on both periods' market potential. This finding corresponds to the results that capacity costs are included both in high-season price and low-season price (the results can be obtained by plugging the optimal capacity, K_t , into high-season and low-season prices).

Price sensitivity affects capacity decision. High-season price sensitivity is negatively related to capacity level. The more price sensitive customers are in high season, the service firm carries the smaller capacity because the firm has can shift elastic high-season demand to low season by reducing low-season price. As cross-period price sensitivity gets larger, demand in one period gets more affected by price change in the other period and customers gets more likely to switch between high and low seasons.

Larger cross-period price sensitivity would allow firms to carry smaller capacity by shifting high-season demand to low-season period. This leads to

Proposition 4.3.1 *Service capacity decreases as cross-period price sensitivity gets larger.*

See the appendix C for proof.

Since services are easier to copy than high-tech products, service market is exposed to new entry and competitive actions. Competition should have a substantial effect on service capacity decision. The model predicts that as service market gets more competitive or services are more substitutable with each other, capacity gets larger. Degree of competition has a positive relationship with optimal capacity. This leads to

Proposition 4.3.2 *Service capacity increases as market gets more competitive.*

See the Appendix C for proof. Although the result seems counter-intuitive, the model suggests that as services market gets more competitive, firms should increase their capacities. The intuition behind it is that as the market gets more competitive, customers get more price-sensitive and service firms would lower their prices so that they can attract more customers to their facilities. As competition gets more severe, firms need build more capacity.

4.4 Services Pricing Decisions

Degree of competition can have a significant impact on optimal high-season and low-season prices. As expected, both high-season price and low-season price are negatively related to degree of competition ($\partial p_h / \partial \gamma < 0$, $\partial p_l / \partial \gamma < 0$ in Equations (4.6) and (4.7)). Increase in market demand in one period would increase both season and off-season prices. In Equations (4.6) and (4.7), upward shift in demand in low-season period would increase both high-season price and low-season price ($\partial p_h / \partial a_l > 0$, $\partial p_l / \partial a_l > 0$). These results are due to the fact that high-season and low-season demand are interdependent and that change in demand given fixed capacity make service mangers adjust prices to match new demand to capacity in each period and to maximize total profits across the two periods.

Demand interdependencies between the two period's result in the effect of cross-period price sensitivity on high-season and low-season prices. As demand gets more price sensitive between the two periods, demand in one period is more sensitive to price change in the other period. This can be caused by customers'greater flexibility in time for service use. For instance, vacationers have greater flexibility in traveling and lodging than business travelers do. Price difference between high season and low season would have larger impact on the choice of timing of service use for vacationers than for business travelers. In Equation (4.6), as cross-period price sensitivity increases, high-season price decreases. In Equation (4.7), as cross-period price sensitivity increases, low-season price also increases.

Proposition 4.4.1 As cross-period price sensitivity increases, high-season price decreases while low-season price increases.

See Appendix C for proof. When cross-period sensitivity is large, high-season service can be a substitute to low-season service. The competition of one service between two periods can reduce high-season price and raise low-season price for maximizing total profits. The cross-period competition allows service firms to carry smaller capacity (see Proposition 4.3.2).

Service firms do not consider only high-season and low-season prices, but also discount depth, difference between the two periods because discount depth has an effect on demand in each period and thus profitability. Discount depth (DD) in the competitive setting is

$$\text{DD}=\frac{[K+(b_h+\gamma)(c+k)](2b_l+\gamma)-[a_l+(b_l+\gamma)c-k\theta](b_h+\gamma)}{(b_h+\gamma+\theta)(2b_l+2\theta+\gamma)-2\theta^2} \tag{4.9}$$

Since own price effect is assumed greater than cross-period price effect, $b_h > \theta$. Discount depth increases as off-season market potential (a_l) gets smaller ($\partial DD/\partial a_l < 0$). Price sensitivity of low-season demand also influences discount depth ($\partial DD/\partial b_l > 0$): the more price sensitive low-season customers get, the larger discount depth should be.

Cross-period price sensitivity also has an effect on discount depth. High-season price is negatively related to cross-period sensitivity and low-season price is positively related to cross-period sensitivity (see Proposition 4.4.2). This leads to

Corollary 4.4.1 As cross-period price sensitivity increases, the discount depth decreases.

See Appendix C for proof. This finding allude to the possibility that price difference between high season and low season can be used as a measure of cross-period price sensitivity, with other things equal.

Degree of competition also affects discount depth: the more competitive the market is, the larger discount depth gets. More competition can offer customers more opportunities to compare prices of competing firms. Thereby competition can be positively related to higher price sensitivity of customers. Thus more competition in the market would expand the discount depth. This leads

Proposition 4.4.2 In a competitive market, discount in low-season should increase, as degree of competition increases.

See Appendix C for proof. The positive effect of competition on discount depth implies that competition reduces low-season price more than high-season price because competition affects negatively both prices ($\partial p_h / \partial \gamma < 0$, $\partial p_l / \partial \gamma < 0$). In low-season period, service firms with empty capacity may compete more fiercely in prices.

Discount depth is positively related to capacity level ($\partial DD / \partial K > 0$). The positive relationship between capacity level and discount depth implies that a service firm with a larger capacity should lower its price to fill more capacity because unoccupied or unsold capacity has no value at all. This leads to

Proposition 4.4.3 As capacity increases, discount depth also increases.

See Appendix C for proof. This finding can be explained by assertion of inventory clearing pricing in retailing. Even though unsold service capacity does not incur inventory- holding costs, unoccupied capacity has no value and thus deteriorates profitability. This can be also compatible with the findings of Tirole (1988) that a large capacity firm should have a Stackleberg-follower profits in equilibrium. Thus a service firm with the larger capacity would be in an unfavorable position for price competition. In sum, with no uncertainty of demand, large-capacity service firms should provide greater discount in low-season period to make an efficient use of their capacities.

4.5 Effects of Demand Fluctuations

In this section, demand is assumed not only to vary seasonally but also to fluctuate randomly within a certain range. Unlike seasonal change in demand, the random fluctuation denotes increase and decrease in demand due to unexpected factors such as change in economic situations (e.g., Gross Domestic Products, unemployment, etc.) and change in customer tastes. However, without any change in economic situation or tastes, customer can arrive at the service facility in a random pattern. In the queuing theory, random customer arrivals can be described by the probability distribution, specifically the time between arrivals (Levin, Rubin and Stinson 1986, p.655).

Capacity decision gets complicated under demand uncertainty. As demand randomly varies, errors may arise due to variations of demand forecasts, model misspecification, etc. Reliability of demand forecasts may affect the variance of the profit function. Thus Service firms should take demand fluctuations into account when they decide capacity level.

Additional Notation

ε_m = random component of demand uniformly distributed, $[-\delta_m, \delta_m]$ and its density is

$f(\varepsilon_m)=1/2\delta_m$

i,j = firm i,j

m,n = time period, i.e., m,n=s(season) or o(off-season)

The distribution of random component is assumed to be uniform and the probability that the random variable takes any value within a certain range is identical. The reason for assumption of uniform distribution is that there is no reason to assume some mass points in random variations in demand a priori.

Demand function is assumed to have two components, deterministic and random components. The demand function for firm i in period m is

Demand

$$D_{im} = a_m - b_m p_{im} + \gamma p_{jm} + \theta p_{in} + \varepsilon_m . \tag{4.10}$$

The random component of demand function denotes demand fluctuations and thereby sales vary in relation to capacity level.

Then the sales under demand uncertainty would be

$$S_i = \begin{cases} a_m - b_m p_{im} + \gamma p_{jm} + \theta p_{in} + \varepsilon_m & \text{if } D \le K_i \text{, i.e., } -\delta_m \le \varepsilon \le K_i - (a_m - b_m p_{im} + \gamma p_{jm} + \theta p_{in}) \\ K_i & \text{if } D > K_i \text{, i.e., } K_i - (a_m - b_m p_{im} + \gamma p_{jm} + \theta p_{in}) < \varepsilon < \delta_m \end{cases} \tag{4.11}$$

The service firm is assumed to be risk neutral and thus expected profit is equivalent to its expected utility. The firm's expected profits are

$$E(\Pi_i)=(p_{jm}-c)\left[\int_{-\delta_m}^{K_i-\mu}(a_m-b_m p_{im}+\gamma p_{jm}+\theta p_{in}+\varepsilon_m)\frac{1}{2\delta_m}d\varepsilon_m+\int_{K_i-\mu}^{\delta_m}K_i\frac{1}{2\delta_m}d\varepsilon_m\right]-kK_i$$

$$\text{where}\quad \mu=a_m-b_m p_{im}+\gamma p_{jm}+\theta p_{in} \tag{4.12}$$

From the expected profit function for firm *i*, we solve for optimal capacity, K_i,

$$K_i=(a_m-b_m p_{im}+\gamma p_{jm}+\theta p_{in}+\delta_m)-\frac{2k\delta_m}{p_{im}-c} \tag{4.13}$$

The first term denotes the largest possible demand. Optimal capacity in the presence of random demand fluctuation can be determined by maximum demand, degree of demand fluctuation, capacity costs and unit margin. Capacity increases as the market potential gets larger. Capacity increases as unit margin gets bigger. Demand may change a lot in one area while demand may stay almost the same around the year in another. Thus it may be very important for service managers to look at the effect of demand fluctuation in their service target areas on capacity decision. The model predicts that optimal capacity decreases as demand fluctuation gets larger. Even though service firms are not risk averse (i.e., we assume firms are risk neutral), they should under-build their capacities in the presence of large demand fluctuation. This leads to

Proposition 4.5. When demand is subject to random variations, service capacity should decrease as demand fluctuation gets larger.

See the Appendix C for proof. Proposition 4.5 implies that when capacity costs are large enough, big demand fluctuation make carrying huge capacity very costly. Services firms that carry large capacities to accommodate sudden augmentation in demand may have much empty capacity most times. Thus in a highly turbulent market, service firms should have low capacities even though they may lose some customers in the presence of a spike in demand.

4.6 Empirical Analysis and Managerial Implications

In order to check the analytical results in a real service situation, secondary data have been collected in the hotel industry. The geographical areas of the data consist of eight regions in the state of Florida. The hotels in the data set are all members of the Florida Hotel and Motel Association (FHMA). The FHMA annually publishes a brochure where high season and low season prices by each hotel are listed. The discount depth is calculated as a difference between season and off-season prices. In case where a hotel offers more than two tiers of prices during a year, the highest price and the lowest price are selected as season and off-season price, respectively. The FHMA also collects data on the number of rooms that can be a measure of capacity. The association also collects data on occupancy rates that is ratio of number of rooms occupied to total number of rooms or capacity of member hotels and motels. Demand fluctuation is measured by variance in occupancy rates.

As one measure of size of market potential, the Bureau of Economic Analysis, the Florida Department of Commerce, collects data on visitors to Florida on a quarterly basis.

Interviews are conducted on a person-to-person basis with Florida's domestic visitors as they are completing their stay and leaving the state. Individuals interviewed meet two criteria: they are not residents of Florida and they have been in the state at least one night and no more than 180 nights. For match to this quarterly visitor data, the season/off-season prices and occupancy rates are transformed to quarterly averages. The visitor study also collects data on the number of properties (as a measure of competition). In sum, the data set is on both cross-sectional and longitudinal basis. Empirical findings are summarized in Tables 4-1, 4-2 and 4-3.

From the empirical results in Table 4-1, hotel capacities in an area depend on both magnitude of visitors in the region and the number of competitors in the same lodging market. As more visitors arrive in a region, hotels should have larger capacity. This supports that optimal capacity is a positive function of size of potential customer base that is positively correlated with the number of visitors.

Table 4-1 Factors Affecting Capacity Decision

Dependent Variables	*Explanatory Variables*	*Parameter Estimates*	*t-value*	*p-value*
Capacity n=32, F(2,29)=17.479				
	# Properties	.658085 (.118603)	5.549	0.0001
	Demand Variance	-5.985612 (1.317238)	-4.544	0.0001

*Standard errors in parentheses

Regarding effect of competition on capacity, as there are more competitive hotels in the area, each hotel should have larger capacity. This supports proposition 4.3.2. The finding is counter intuitive and intriguing since conventional wisdom posits that increase in the number of competitors in the same market decrease sales for each firm and that each hotel should reduce capacity with competition. In the hotel industry, however, as more competitive hotels exist in the market, firms would decrease prices and attract more customers to their facilities and thus need larger capacities. Demand fluctuations that are measured by occupancy rate variances have a negative effect on capacity. This supports Proposition 4.5. From a managerial standpoint, out of possible explanatory variables for optimal capacity, significant variables are very important because capacity requires huge amount of capital, and it also takes much time so not flexible in the short run. Based on the analytical and empirical results, capacity decision should not only take into account trend of the number of visitors, but also how many hotels compete for customers in the same region. According to the proposition 4.5, degree of demand variance in the area should be also considered. As demand fluctuations get larger, hotels have smaller capacities in the data. This shows that hotels are risk averse with uncertainty of demand.

Once the capacity level is set, hotels and motels may be concerned with how they charge high-season rates and low-season rates. In the Table 4-2, the number of rooms influences positively high-season rates. The number of rooms also affect positively low-season rates.

Table 4-2 Effects of Capacity Level on Rates

Dependent Variables	*Explanatory Variables*	*Parameter Estimates*	*t-value*	*p-value*
High-Season Rate N=19, F(1,17)=16.304	# Rooms	0.137062 (0.033944)*	4.038	0.0009
Low Season Rate n=19, F(1,17)=19.003	# Rooms	0.114215 (0.0262006)	4.359	0.0004

*Standard errors in parenthesis

These results also suggest that customers perceive large capacity hotels as high-quality ones and higher reservation prices.

Table 4-3 summarizes the empirical findings on the factors affecting discount depth. Discount depth also depends on analysis of competitors. Even though each hotel has unique features to some degree, increase in the number of competitive hotels in the same market would lead to larger discounts offered in low-season period. The number of competitive hotels has a positive effect on discount depth. This finding supports Proposition 4.4.2. From the empirical results, discount depth between high season and low season should be driven by the degree of competition. As competition gets severer, customers get more price- sensitive and are more likely to compare rates of competitive hotels. Competition leads to larger discount depth. This can be compared with the findings in the peak load pricing literature that a utility company should charge lower price in off-peak to shift excess demand from peak to off-peak. Unlike utilities, however, hotels face severe competitions and may have low demand for the most time of the year except for a short period of high demand. Thus hoteliers should consider competitive actions when they make decision on depth of discount offered in low-season period.

Table 4-3 Factors Affecting Discount Depth

Dependent Variables	*Explanatory Variables*	*Parameter Estimates*	*t-value*	*p-value*
Discount Depth N=32, F(1,30)=4.271	# Properties	0.000796 (0.000385)*	2.067	0.0475
n=19, F(1,17)=7.391	# Rooms	0.051787 (0.01904899)	2.719	0.0146

*Standard errors in parenthesis

For checking the effect of capacity level on discount depth, we use data on the number of rooms in hotels and differences between high-season price and low-season price in Panama City Beach Area, Florida. The empirical results in the Table 4-3 show that discount depth increases as the number of rooms gets larger. This supports Proposition 4.4.3 that capacity has a positive effect on discount depth.

4.7 Summary and Directions for Future Research

In the paper, we propose a comprehensive model of a two-stage capacity and pricing decision in services marketing. In the model, we consider demand seasonality and random demand fluctuation. The two-stage sequential model predicts that capacity decision should depend on degree of competition, cross-period price sensitivity and magnitude of demand fluctuation. As the service market gets more competitive, firms should have larger capacities. As service demand fluctuates more, firms should have smaller capacities. Regarding pricing decision, the proposed model predicts that service pricing should depend on capacity level, degree of competition and cross-period price sensitivity. The capacity level positively affects discount depth between high season and low season. Discount depth between the two periods increases as degree of competition increases. As cross-period price sensitivity increases, high-season price decreases while low-season price increases. Thus as cross-period price sensitivity increases, discount depth decreases.

These analytical results are checked by the cross-sectional data collected in the Florida hotel industry. The number of hotel properties is used as a measure of competition and occupancy variance as a measure of demand fluctuations. Hotel capacity is a positive function of the number of properties. The capacity is a negative function of demand variance. The capacity level of hotels has a positive effect on the difference between high-season price and low-season price. The number of properties in each region also has a positive effect on discount depth.

An immediate extension of the paper is to investigate the optimal number of price switch points during a cycle of demand seasonality. The research question is when service firms adopt uniform pricing versus multiple-tier pricing or peak load pricing and how many tiers of prices are optimal under a certain condition of demand and costs. Since it is easy to observe that in one area, some hotels provide single rates all year round while others provide multiple tiers of rates, it would be interesting to investigate the conditions for uniform pricing versus multiple pricing.

Another interesting area of research is to analyze the mechanism in which customers choose the timing of service consumption. Since service demand varies seasonally in a predictable pattern, there are periods of being crowded and being not crowded given a certain level of capacity. Non-storability of services makes service mangers take into account consumption timing for planning service productions. It is expected that customer characteristics such as cost of waiting time and degree of sociability as well as managerial decisions such as time-dependent pricing and reservations system can affect choice of timing. It is also intriguing to examine signaling of crowd in the service facility; some people may view crowd as inconvenience and low price while others may regard it as a signal of popularity and high quality. Signaling theory can be applied to the research question on crowd to provide service managers with insights on how they manage crowd in service market positioning.

APPENDIX A

SERVICE PRICING EXAMPLES

Table A-1. Pricing Examples in Service Businesses

RESTAURANT

		Regular Dinner	Early Special Dinner
Entrees	Top Sirloin(small)	$ 8.95	$ 8.49
	Top Sirloin(large)	10.95	9.99
	Prime Rib (regular)	11.45	9.99
	Kensington Club	11.95	10.99
	Stuffed Flounder Maître D'	9.95	8.79
	Blackened Snapper	9.95	8.99
	Southwest Chicken	9.95	7.29
	Chicken Pasta	8.95	7.79
Desserts	Chocolate Fudge	2.25	Free
	Strawberry Cake	2.25	Free
	Yogurt	2.25	Free
Beverage	Soda	1.33	Free
	Coffee	1.33	Free
	Tea	1.33	Free

BAR

	Regular Hours	Happy Hours
Draft Beer	$1.5	$0.99
Wine	Regular price	2 for 1(or 3 for 1)
Beer	Regular price	2 for 1(or 3 for 1)
Liquor	Regular price	2 for 1(or 3 for 1

MOVIE THEATER

	before 6 pm	after 6 pm
Adults	$2.50	$5.00
Children	2.50	3.50

GOLF CLUB

	before 1 pm	1 pm - 4 pm	after 4 pm
Weekdays	$18	$15	$11
Weekends	22	16	11

CAR RENTAL Co.

Size	Weekdays (Mon. pm- Thurs. am)	Weekends (Thurs. pm- Mon. am)
Premium	$ 40.99	$40.99
Full size	37.00	36.00
Mid Size	33.00	32.00
Compact	31.00	30.00

APPENDIX B

PROOFS OF PROPOSITIONS IN CHAPTER TWO

Proof of Proposition 2.1

From P_{p1}^* and P_{p2}^*, we get the following relative markups for service one and service two in peak period. In the case of complements,

$$\frac{P_{p1}^* - c_{p1}}{P_{p1}^*} = \frac{\lambda_1}{P_{p1}^*} + \frac{1}{\varepsilon_{p1}} - \frac{1}{\varepsilon_{o1,p1}} + \frac{1}{\varepsilon_{p2,p1}} - \frac{1}{\varepsilon_{o2,p1}} \text{ and}$$

$$\frac{P_{p2}^* - c_{p2}}{P_{p2}^*} = \frac{1}{\varepsilon_{p2}} - \frac{1}{\varepsilon_{o2,p2}} + \frac{1}{\varepsilon_{p1,p2}} - \frac{1}{\varepsilon_{o1,p2}}$$

Let's assume same own price elasticities ($\varepsilon_{p1} = \varepsilon_{p2}$) and same corresponding cross price elasticities ($\varepsilon_{o1,p1} = \varepsilon_{o2,p2}$, $\varepsilon_{p2,p1} = \varepsilon_{p1,p2}$ and $\varepsilon_{o2,p1} = \varepsilon_{o1,p2}$).

Then $\dfrac{P_{p1}^* - c_{p1}}{P_{p1}^*}$ - $\dfrac{P_{p2}^* - c_{p2}}{P_{p2}^*} = \dfrac{\lambda_1}{P_{p1}^*} > 0$ as required.

In case of substitute services, the proposition 1 can be similarly proved.

$$\frac{P_{p1}^* - c_{p1}}{P_{p1}^*} = \frac{\lambda_1}{P_{p1}^*} + \frac{1}{\varepsilon_{p1}} - \frac{1}{\varepsilon_{o1,p1}} - \frac{1}{\varepsilon_{p2,p1}} + \frac{1}{\varepsilon_{o2,p1}} \text{ and}$$

$$\frac{P_{p2}^* - c_{p2}}{P_{p2}^*} = \frac{1}{\varepsilon_{p2}} - \frac{1}{\varepsilon_{o2,p2}} - \frac{1}{\varepsilon_{p1,p2}} + \frac{1}{\varepsilon_{o1,p2}}.$$

Based on the assumptions of same own, cross-period and cross-firm elasticities,

$$\frac{P_{p1}^{*}-c_{p1}}{P_{p1}^{*}} - \frac{P_{p2}^{*}-c_{p2}}{P_{p2}^{*}} = \frac{\lambda_1}{P_{p1}^{*}} > 0$$ as required.

Proof of Proposition 2.2

We can prove proposition 2.2 by doing the following comparative statics. Differentiating d_1, the difference between peak price and off-peak price for service 1, with respect to ε_{p1} , own price elasticity of peak demand for the service, reveals

$$\frac{\partial d_1}{\partial \varepsilon_{p1}} = \frac{-\dfrac{c_{p1}+\lambda_1}{\varepsilon_{p1}^{2}}}{\left(1-\dfrac{1}{\varepsilon_{p1}}+\dfrac{1}{\varepsilon_{o1,p1}}-\dfrac{1}{\varepsilon_{p2,p1}}+\dfrac{1}{\varepsilon_{o2,p1}}\right)^{2}} < 0$$

Differentiating d_1 with respect to ε_{o1} , own price elasticity of off-peak demand for the same service, reveals

$$\frac{\partial d_1}{\partial \varepsilon_{o1}} = \frac{-\dfrac{c_{o1}+\lambda_2}{\varepsilon_{o1}^{2}}}{\left(1-\dfrac{1}{\varepsilon_{o1}}+\dfrac{1}{\varepsilon_{p1,o1}}-\dfrac{1}{\varepsilon_{p2,o1}}+\dfrac{1}{\varepsilon_{o2,o1}}\right)^{2}} < 0$$

The above two partial derivatives show that discout depth in off-peak period increases as own price elasticity in peak increases and that discout depth in off-peak period decreases as own price elasticity in off-peak increases. Hence the proposition 2.2 is proved.

APPENDIX C

PROOF OF PROPOSITIONS IN CHAPTER 4

Proof of Proposition 4.3.1

In the capacity decision of the two stage sequential model, cross-period price sensitivity is measured by θ. Differentiating the optimal capacity in the equation (4.8) with respect to θ, we get $\frac{\partial K^*}{\partial \theta} < 0$ as required.

Proof of Proposition 4.3.2

In the model of capacity decision, degree of competition or substitutibility is measured by γ. Differentiating the optimal capacity in the equation (4.8) with respect to γ, we get $\frac{\partial K^*}{\partial \gamma} > 0$ as required.

Proof of Proposition 4.4.1

In order to check the effect of cross-period price sensitivity on high-season price and low-season price, we differentiate with respect to cross-period price sensitivity, θ, the optimal high-seson price in the equation (4.6) and the low-season price in the equation (4.7). Then $\frac{\partial p_h^*}{\partial \theta} < 0$ and $\frac{\partial p_l^*}{\partial \theta} > 0$ as required.

Proof of Corollary 4.4.1

Differentiating the discount depth (DD) in the equation (4.9), we get $\frac{\partial DD^*}{\partial \theta} > 0$ as required.

Proof of Proposition 4.4.2

Differentiating the discount depth (DD) in the equation (4.9), we get $\frac{\partial DD^*}{\partial \gamma} > 0$ as required. Degree of competition is negatively related with associated high-season and low-season prices $\left(\frac{\partial p_h^*}{\partial \gamma} < 0, \frac{\partial p_l^*}{\partial \gamma} < 0 \right)$. Positive effect of competition on discount depth is due to the results that competition effect is greater on low-season price than on high-sesaon price.

Proof of Proposition 4.4.3

Differentiating the discount depth (DD) in the equation (4.9) with respect to capacity level K, we get $\frac{\partial DD^*}{\partial K} > 0$ as required.

Proof of Proposition 4.5

In the stochastic demand model, degree of demand fluctuations is measured by the range of values that the random component of demand (ε_m). For investigating the effect of degree of demand fluctuation on capacity decision, we do the comparative statics by differentiating the optimal capacity with respect to range of demand fluctuations, δ_m.

$$\frac{\partial K^*}{\partial \delta_m} = 1 - \frac{2k}{p_{tm} - c} < 0 \qquad \text{if } k > (p_{tm} - c) / 2 .$$

This side condition should be met because capacity costs may be much larger than a half of unit margin. Even though we assume a uniform distribution for random component of the demand function, it is possible to show the above results by using a general probability density function. In that case, degree of demand fluctuation can be neasured by variance of random demand.

REFERENCES

Bailey, E. and L. White (1974) "Reversals in Peak and Off-peak Prices, " *Bell Journal of Economics*, Vol. 5,1: 75-92

Banerjee, A. and L. Summers (1987) "On Frequent Flyer Programs and Other Loyalty-Inducing Arrangements," Harvard Institute of Economic Research Discussion Paper No.1337.

Becker, Gary S. (1991) " A Note on Restaurant Pricing and Other Examples of Social Influences on Price," *Journal of Political Economy*, Vol. 99: 1109-1116.

Beckman, M. (1965) "Edgeworth-Bertrand Duopoly Revisited," R. Henn, ed., *Operations Research Verfahren III* Meisenhein: Verlag Anton Hain.

Benoit, J. and V. Krishna (1987) "Dynamic Duopoly: Price and Quantities," *Review of Economic Studies*, Vol.54: 23-35.

Berg, S. and J. Tschirhart (1988) *Natural Monopoly Regulation*, Cambridge: Cambridge University Press.

Bergstrom, T. and J. Mackie-Mason (1991) " Some Simple Analytics of Peak-Load Pricing," *Rand Journal of Economics*, Vol. 22: 241-249.

Bryant, J. (1990) " Competitive Equilibrium with Price Setting Firms and Stochastic Demand," *International Economic Review*, Vol. 21: 619-626.

Burness, H. and R. Patrick (1991) "Peak-load Pricing with Continuous and Interdependent Demand," *Journal of Regulatory Economics*, Vol.3: 69-87.

Bulow, J., J. Geanakopolos and P. Klemperer (1985) "Holding Idle Capacity to Deter Entry," *The Economic Journal*, Vol. 95: 178-182.

Caminal, R. and C. Matutes (1990) "Endogenous Swithching Costs in a Duopoly Model," *International Journal of Industrial Organization*, Vol.8: 353-373.

Carlson, J. and D. Pescatrice (1980) "Persistent Price Distributions," *Journal of Economics and Business*, Vol. 33: 21-27.

Carlton, D. (1978) "Market Behavior with Demand Uncertainty and Price Inflexibility," *The American Economic Review*, Vol. 68: 571-587.

Couglan, A. and M. Mantrala (1992) "Dynamic Competitive Pricing Strategies," *International Journal of Research in Marketing*, Vol. 9: 91-108.

Crew, M. and P. Kleindorfer (1986) *The Economics of Public Utility Regulation*, The MIT Press, Cambridge, Massachusetts.

Dahlby, B. and West (1986) " Price Dispersion in an Automobile Insurance Market *Journal of Political Economy*, Vol. 94: 418-439.

Darby, M. and E. Karni (1973) "Free Competition and the Optimal Amount of Fraud," *Journal of Law and Economics*, Vol.16: 67-88.

Davidson, C. (1988) "Equilibrium in Servicing Industries: An Economic Application of Queuing Theory, *Journal of Business*, Vol. 61: 347-367.

Davidson, C. and R. Deneckere (1986) "Long-Run Competition in Capacity, Short-Run Competition in Price and the Cournot Model," *Rand Journal of Economics*, Vol. 17: 404-415.

Davis, M. and T. Vollman (1990) "A Framework for Relating Waiting Time and Customer Satisfaction in a Service Operation,"*Journal of Services Marketing*, Vol. 4: 61-70.

De Vany, A. (1976) "Uncertainty, Waiting Time, and Capacity Utilization: A Stochastic Theory of Product Quality," *Journal of Political Economy*, Vol.84: 523-540.

Deneckere, R and J. Peck (1995) "Competition over Price and Service Rate when Demand is Stochastic: A Strategic Analysis, *Rand Journal of Economics*, Vol. 26: 148-162.

Delta Airlines (1996) Dream Vacations brochure -- Florida Beaches.

Eaton, B. and R. Lipsey (1977) "The Theory of Market Preemption: The Persistence of Excess Capacity and Monopoly in Growing Spatial Markets," *Economica*, Vol. 46:
149-158.

Fisher, R. (1991) "Durable Differentiation Strategies for Services," *Journal of Services Marketing*, Vol. 5: 19-28.

Friedman, J. (1988) "On the Strategic Importance of Prices versus Quantities," *Rand Journal of Economics*, Vol.19: 607-622.

Gallant, A. R. and R. Koenker (1984) "Costs and Benefits of Peak Load Pricing of Electricity: A Continuous-time Econometric Approach," *Journal of Econometrics*, Vol. 26: 83-113.

Gelman, J. and S. Salop (1983) "Judo Economics: Capacity Limitation and Coupon Competition,"*Bell Journal of Economics*, Vol. 14: 315-325.

Gerstner, E. (1986) "Peak Load Pricing in Competitive Markets," *Economic Inquiry*, Vol. 24: 349-361.

Gertner, R. (1985) " Simultaneous Move Price-Quality Games and Non-Market Clearing Equilibrum," Mimeo, Massachusetts Institute of Technology, Cambridge.

Gould, J. (1978) " Inventories and Stochastic Demand: Equilibrium Models of the Firm and Industry, *Journal of Business*, Vol. 51: 1-42.

Hands, D. Wade (1991) *Introductory Mathematical Economics*, D.C. Heath and Company, Lexington, Massachusetts.

Harsanyi, J. (1973) "Games with Randomly Distributed Payoffs: A New Rational for Mixed Strategy Equilibrium Points," *International Journal of Game Theory*, Vol.2: 1-23.

Haynes, P. (1990) "Hating to Wait: Managing the Final Service Encounter,"*Journal of Services Marketing*, Vol. 4: 20-26.

Herk, L. (1993) "Consumer Choice and Cournot Behavior in Capacity-Constrained Duopoly Competition, *Rand Journal of Economics*, Vol. 24: 399-417.

Hirshleifer, Jack (1958) "Peak Loads and Efficient Pricing: Comment," *Quarterly Journal of Economics*, Vol. 72: 451-462.

Jones, P. and D. Hamilton (1992), "Yield Management: Putting People in the Big Picture,"*The Cornell Hotel and Restaurant Administration Quarterly*, Vol.33: 89-95

Klemperer, P (1987 a) "Markets with Consumer Switching Costs, *Quarterly Journal of Economics*, Vol.102: 375-394.

Klemperer, P. (1987 b) "The Competitiveness of Markets with Switching Costs," *Rand Journal of Economics*, Vol.18: 138-150.

Kreps, D. and J. Scheinkman (1983) "Quantity Pre-commitment and Bertrand Competition Yield Cournot Outcomes," *The Bell Journal of Economics*, Vol.14: 326-337.

Kreps, D. (1990) *A Course in Microeconomic Theory*, Princeton Univerisity Press, Princeton.

Levin, R., D. Rubin and J. Stinson (1986) *Quantitative Approaches to Management*, McGraw-Hill Inc., New York.

Levitan, R. and M. Schubik(1972) "Price Duopoly and Caapcity Constraints," *International Economic Review*, Vol. 13: 111-122.

Lieberman, M. (1987) "Excess Capacity as a Barrier to Entry: An Empirical Appraisal," *The Journal of Industrial Economics*, Vol. 35: 607-627.

Lieberman, M. (1989) "Capital Utilization: Theoretical Models and Empirical Tests *European Journal of Operations Research*, Vol. 40.

Lippman, S. and K. McCardle (1997) "The Competitive Newsboy,"*Operations Research*, Vol. 45: 54-65.

Mantrala, M.K. and K. Raman (1996) "The Multi-store Retailer's Style-goods Order Quantity Problem and Supplier's Optimal Returns Policy," University of Florida Working Paper.

McGuire, T. and R. Staelin (1983) "An Industry Equilibrium Analysis of Downstream Vertical Integration," *Marketing Science*, Vol. 2: 161-191.

Milgrom P. and J. Roberts (1982) "Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis," *Econometrica*, Vol. 50: 443-460.

Moorthy, K. (1985) "Using Game Theory to Model Competition," *Journal of Marketing Research*, Vol. 32: 262-282.

Nagle, T. (1984) "Economic Foundations for Pricing," *The Journal of Business*, Vol. 57: S3-26.

Nalebuff, B. and J. Riley (1985) "Asymmetric Equilibria in the War of Attrition," *Journal of Theoretical Biology*, Vol.113: 517-527.

Osborne, M. and C. Pitchik (1986) " Price Competition in a Capacity-Constrained Duopoly," *Journal of Economic Theory*, Vol. 38: 238-260.

Overstreet, G., Jr. (1993) " Creating Value in Oversupplied Markets: The Case of Charlottesville, Virginia, Hotels," *The Cornell Hotel Restaurant Administration Quarterly*, Vol. 34: 68-96.

Paraskevopoulos, D., E. Karakitsos and B. Rustem (1991) "Robust Capacity Planning Under Uncertainty,"*Management Science*, Vol. 37: 787-800.

Peters, M. (1984) "Bertrand Equilibrium with Capacity Constraints and Restricted Mobility,"*Econometrica*, Vol. 52: 1117-1127.

Platt, A. (1981) "Traditional Peak Load Pricing Theory: A Synthesis,"*Journal of Economic Studies*, Vol. 8: 47-51.

Png, I.P.L. (1989) "Reservations: Customer Insurance in the Marketing of Capacity," *Marketing Science*, Vol. 8: 248-264.

Pratt, J., D. Wise and R. Zeckhauser (1979) "Price Differences in Almost Competitive Markets,"*Quarterly Journal of Economics*, Vol. 93: 189-211.

Pressman, I. (1970) "A Mathematical Formulation of Peak- Load Pricing Problem," *Bell Journal of Economics*, Vol.1: 304-326.

Rhee, H. and S. Shugan (1997) "Pricing Strategies for Complementary Services," University of Florida Working Paper.

Sasser, W. (1976) "Match Supply and Demand in Service Industry," *Harvard Business Review*, Vol.54: 133-140.

Sheel, A. (1995) "Monte Carlo Simulations and Scenario Analysis - Decision-Making Tools for Hoteliers," *The Cornell Hotel and Restaurant Administration Quarterly*, Vol.36: 18-26.

Shugan, S., *Services Strategies: A Contextual Approach*, Richard D. Irwin Inc., Homewood, Illinois, forthcoming.

Shugan, S. (1989) "Product Assortment In A Tripoly," *Management Science*, Vol.35: 304-320.

Steiner, P. (1957) "Peak Loads and Efficient Pricing," *Quarterly Journal of Economics*, Vol. 71: 585-610.

Stigler, G. (1961) "The Economics of Information," *Journal of Political Economy*, Vol. 69: 213-225.

Tirole, J. (1988) *The Theory of Industrial Organization*, MIT Press, Cambridge.

Varian, H. (1988) "A Model of Sales," *The American Economic Review*, Vol. 70: 651-659.

Venkatesh, R. and V. Mahajan (1993) "A Probabilistic Approach to Pricing a Bundle of Products or Services," *Journal of Marketing Research*, Vol. 30: 494-508.

von Weizsacker, C. (1984) "The Costs of Substitution," *Econometrica*, Vol. 52: 1085-1116.

Weitz, B. (1985) "Introduction to Special Issues on Competition in Marketing," *Journal of Marketing Research*, Vol.32: 229-236.

Wernerfelt, B. (1991) " Brand Loyalty and Market Equilibrium,"*Marketing Science*, Vol.10: 229-245.

Williamson, O. (1966) "Peak-Load Pricing and Optimal Capacity Under Indivisibility Constraints," *American Economic Review*, Vol. 55: 810-827.

Wolff, C. (1997) "Keeping the Boom on the Course," *Lodging Hospitality*, Vol. 53: 36-39.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Steven M. Shugan, Chair
Russell Berrie Eminent Scholar in
Marketing

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Barton A. Weitz
J.C. Penney Eminent Scholar in Retail
Management

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

M. Murali Krishna
Murali K. Mantrala
Associate Professor of Marketing

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

JMahajan
Jayashree Mahajan
Assistant Professor of Marketing

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Sanford Berg
Professor of Economics

This dissertation was submitted to the Graduate Faculty of the Department of Marketing in the Warrington College of Business Administration and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December, 1997

Richard

Dean, Graduate School